



Instituto Nacional de Estadística

OPOSICIONES AL CUERPO DE DIPLOMADOS EN ESTADÍSTICA DEL ESTADO

BOE NÚM. 270, DE 12 DE OCTUBRE DE 2020, PÁG. 87165

Estadística

Índice general

| | |
|---|-----------|
| 1 Fenómenos aleatorios. Conceptos de probabilidad. Propiedades. Independencia de sucesos. Teorema de Bayes. | 1 |
| 1.1 Historia de la probabilidad | 1 |
| 1.2 Interpretaciones de la probabilidad | 1 |
| 1.2.1 La interpretación frecuentista de la probabilidad | 1 |
| 1.2.2 La interpretación clásica de la probabilidad | 2 |
| 1.2.3 La interpretación subjetiva de la probabilidad | 3 |
| 1.3 Experimentos y sucesos | 3 |
| 1.4 Teoría de conjuntos | 4 |
| 1.4.1 Espacio muestral | 4 |
| 1.4.2 Relaciones de la teoría de conjuntos | 4 |
| 1.4.3 El conjunto vacío | 5 |
| 1.4.4 Operaciones de la teoría de conjuntos | 5 |
| 1.5 Definición axiomática de probabilidad | 7 |
| 1.5.1 Consecuencias de la definición de probabilidad | 8 |
| 1.6 Probabilidad condicional | 10 |
| 1.7 Sucesos independientes | 11 |
| 1.7.1 Ley multiplicativa para probabilidades condicionales | 16 |
| 1.8 Teorema de Bayes | 17 |
| 1.8.1 Probabilidad y particiones | 17 |
| 2 Variables aleatorias. Variables discretas. Función de probabilidad. Variables continuas. Función de densidad. Propiedades. | 21 |
| 2.1 Variables aleatorias. | 21 |
| 2.2 Función de distribución de una variable aleatoria. | 22 |
| 2.2.1 Propiedades de la función de distribución. | 23 |
| 2.3 Modelos de probabilidad más relevantes | 27 |
| 2.3.1 Distribucion de Bernoulli | 27 |
| 2.3.2 Distribucion de Binomial | 28 |
| 2.3.3 Distribución geométrica | 29 |
| 2.3.4 Distribución binomial negativa | 30 |
| 2.3.5 Distribución Uniforme Discreta | 31 |
| 2.3.6 Distribucion de Poisson | 31 |
| 2.3.7 Distribución Uniforme Continua | 33 |
| 2.3.8 Distribución Exponencial | 34 |
| 2.3.9 Distribución Gamma | 36 |
| 2.3.10 Distribución Normal | 37 |
| 2.3.11 Distribuciones Beta | 40 |
| 3 Esperanza matemática. Propiedades. Varianza. Propiedades. Función característica y función generatriz de momentos. Acotación de Tchebychev | 42 |

| | | |
|----------|---|------------|
| 3.1 | Esperanza matemática. | 42 |
| 3.1.1 | Esperanza de una distribución discreta | 42 |
| 3.1.2 | Esperanza de una distribución continua | 44 |
| 3.2 | Esperanza de una función | 46 |
| 3.3 | Propiedades de la Esperanza | 47 |
| 3.4 | Varianza | 49 |
| 3.4.1 | Propiedades de la varianza | 50 |
| 3.5 | Momentos | 52 |
| 3.5.1 | Función generatriz de momentos | 52 |
| 3.6 | Función Característica | 55 |
| 3.7 | Desigualdades de Markov y Chebyshev | 58 |
| 3.7.1 | Desigualdad de Markov | 58 |
| 3.7.2 | Desigualdad de Chebyshev: | 59 |
| 4 | Distribuciones de varias variables aleatorias. Distribuciones conjuntas y marginales. Independencia entre variables aleatorias. Ejemplos | 62 |
| 4.1 | Función de distribución de un vector aleatorio bidimensional. | 63 |
| 4.1.1 | Propiedades de la función de distribución. | 64 |
| 4.2 | Distribuciones condicionales | 73 |
| 4.3 | Independencia de variables aleatorias | 79 |
| 4.4 | Generalización al caso n-dimensional | 81 |
| 5 | Esperanza de vectores aleatorios. Esperanza de sumas y productos de variables aleatorias. Covarianza. Correlación. Transformaciones lineales de variables aleatorias. | 88 |
| 5.1 | Propiedades de la Esperanza Matemática | 89 |
| 5.2 | Momentos de un vector bidimensional respecto a un punto, respecto al origen y respecto a la media | 90 |
| 5.3 | Correlación lineal entre variables | 94 |
| 5.3.1 | Propiedades del coeficiente de correlación | 95 |
| 5.4 | Cambio de variable | 98 |
| 6 | Propiedades de una muestra aleatoria. Conceptos básicos de muestras aleatorias. Sumas de variables aleatorias de una muestra aleatoria. Muestreo de una distribución normal: propiedades de la media y varianza muestrales y las distribuciones t de Student y F de Snedecor. Estadísticos de orden. | 110 |
| 6.1 | Modelo paramétrico y modelo no paramétrico | 110 |
| 6.2 | Conceptos básicos sobre inferencia. | 111 |
| 6.3 | Estadísticos muestrales. Propiedades | 112 |
| 6.4 | Reproductividad de distribuciones | 116 |
| 6.5 | Distribuciones asociadas al muestreo | 119 |
| 6.6 | Teorema de Fisher. | 127 |
| 6.6.1 | Consecuencias del Teorema de Fisher. | 127 |
| 6.7 | El estadístico de orden: | 128 |
| 7 | Principios de reducción de datos. Introducción. El principio de suficiencia. El principio de verosimilitud: la función de verosimilitud y el principio formal de verosi- | |

| | |
|---|------------|
| militud. | 131 |
| 7.1 Reducción de datos. Introducción | 131 |
| 7.1.1 Principio de suficiencia. | 131 |
| 7.1.2 Criterio de factorización | 132 |
| 7.2 Estadísticos conjuntamente suficientes | 137 |
| 7.2.1 Estadísticos suficientes minimales | 141 |
| 7.3 Función de verosimilitud | 142 |
| 8 Estimación puntual I. Introducción. Métodos para encontrar estimadores: método de los momentos, estimadores máximo-verosímiles, estimadores invariantes. | 147 |
| 8.1 Estimación Puntual | 147 |
| 8.1.1 Propiedades de los estimadores | 147 |
| 8.2 Estimadores invariantes. | 148 |
| 8.3 Estimación máximo verosímil | 149 |
| 8.3.1 Invarianza del estimador máximo verosímil. | 157 |
| 8.3.2 Estimación máximo verosímil en el caso de varios parámetros desconocidos | 158 |
| 8.4 Consistencia de un estimador | 160 |
| 8.5 Estimación por el método de los momentos. | 161 |
| 9 Estimación puntual II. Métodos para evaluar estimadores: error cuadrático medio, suficiencia e insesgadez. | 167 |
| 9.1 Estimadores insesgados o centrados | 167 |
| 9.1.1 Eficiencia | 168 |
| 9.1.2 Cota de Cramér-Rao | 168 |
| 9.1.3 Estadístico Suficiente | 173 |
| 9.1.4 Relaciones entre estimadores. | 184 |
| 10 Tests de hipótesis | 1 |
| 10.1 Tests de hipótesis | 1 |
| 10.1.1 Introducción | 1 |
| 10.1.2 Motivación | 1 |
| 10.2 Planteamiento clásico del contraste de hipótesis | 2 |
| 10.3 Tipos de hipótesis | 3 |
| 10.4 Contrastes de hipótesis simples | 5 |
| 10.4.1 Determinación de regiones críticas óptimas en contrastes de hipótesis simples | 6 |
| 10.5 Contrastes de hipótesis unilaterales | 13 |
| 10.6 Test de hipótesis compuestas (no unilaterales). | 21 |
| 10.6.1 Test de la razón de verosimilitudes | 22 |
| 10.6.2 Test de razón de verosimilitudes en caso de normalidad | 23 |
| 10.7 Interpretación de los resultados proporcionados por los paquetes estadísticos. Concepto de p -valor | 26 |
| 11 Estimación por intervalos. | 1 |
| 11.1 Intervalos de confianza. Introducción | 1 |
| 11.2 Método de la función pivote o cantidad pivotal | 1 |

| | | |
|-----------|---|----------|
| 11.2.1 | Intervalo de confianza para la media de una distribución normal con varianza conocida | 2 |
| 11.2.2 | Intervalo de confianza para la media de una normal con varianza desconocida | 4 |
| 11.2.3 | Intervalo de confianza para la varianza de una normal conocida la media | 6 |
| 11.2.4 | Intervalo de confianza para la varianza de una normal. Media desconocida | 7 |
| 11.2.5 | Intervalo de confianza para la diferencia de medias en poblaciones normales. Varianzas conocidas | 8 |
| 11.2.6 | Intervalo de confianza para la diferencia de medias en poblaciones normales con varianzas desconocidas pero iguales | 9 |
| 11.2.7 | Intervalo de confianza para el cociente de varianzas en poblaciones normales | 10 |
| 11.3 | Método general o de Neyman | 12 |
| 11.4 | Intervalos para parámetros de distribuciones discretas | 15 |
| 11.4.1 | Intervalo de confianza aproximado para p en una población $B(1, p)$ si n es grande ($n \geq 30$). | 15 |
| 11.4.2 | Intervalo de confianza aproximado para λ en el modelo de Poisson | 16 |
| 11.4.3 | Intervalo de confianza para la diferencia de proporciones | 17 |
| 12 | Estadística Descriptiva I. | 1 |
| 12.1 | Concepto de estadística | 1 |
| 12.2 | Las unidades estadísticas | 1 |
| 12.3 | Variables | 2 |
| 12.3.1 | Variables cualitativas y cuantitativas | 3 |
| 12.3.2 | Variables discretas y continuas | 4 |
| 12.3.3 | Escalas | 4 |
| 12.3.4 | Datos agrupados | 5 |
| 12.4 | Distribuciones unidimensionales de frecuencias | 7 |
| 12.4.1 | Frecuencias absolutas y frecuencias relativas | 7 |
| 12.5 | Distribuciones acumuladas | 10 |
| 12.5.1 | Función de distribución para variables ordinales | 11 |
| 12.5.2 | Función de distribución para variables continuas | 13 |
| 12.6 | Tablas estadísticas | 15 |
| 12.7 | Representación gráfica | 16 |
| 12.7.1 | Gráfico de barras | 16 |
| 12.7.2 | Gráfico de tarta | 17 |
| 12.7.3 | Histograma | 19 |
| 12.7.4 | Estimación Kernel de la densidad | 21 |
| 12.8 | Algunas observaciones importantes | 22 |
| 13 | Estadística descriptiva II. | 1 |
| 13.1 | Estadística descriptiva II | 1 |
| 13.2 | Medidas de síntesis de una distribución de frecuencias | 1 |
| 13.2.1 | Distribución de frecuencias de variables cualitativas | 2 |
| 13.2.2 | Distribución de frecuencias de variables cuantitativas | 3 |
| 13.3 | Medidas de posición | 4 |
| 13.3.1 | Media aritmética | 5 |

| | | |
|-----------|---|-----------|
| 13.3.2 | Media geométrica | 9 |
| 13.3.3 | Media armónica | 13 |
| 14 | Estadística descriptiva III | |
| 14.1 | Introducción | 1 |
| 14.2 | Mediana | 2 |
| 14.2.1 | Distribuciones de frecuencias unitarias: número de observaciones impares | 3 |
| 14.2.2 | Distribuciones de frecuencias unitarias: número de observaciones pares | 4 |
| 14.2.3 | Distribuciones de frecuencias no unitarias | 5 |
| 14.2.4 | Distribución de frecuencias agrupadas | 7 |
| 14.3 | Moda | 9 |
| 14.3.1 | Distribuciones de frecuencias de valores sin agrupar | 10 |
| 14.3.2 | Distribuciones de frecuencias de valores agrupados | 12 |
| 14.4 | Cuantiles | 17 |
| 14.4.1 | Cuartiles | 18 |
| 14.4.2 | Deciles | 20 |
| 14.4.3 | Percentiles | 23 |
| | Bibliografía | 26 |
| 15 | Estadística descriptiva IV | 27 |
| 15.1 | Introducción | 27 |
| 15.2 | Recorrido | 29 |
| 15.3 | Varianza | 30 |
| 15.3.1 | Desarrollo del cuadrado de la fórmula de la varianza | 40 |
| 15.3.2 | Cuasivarianza | 40 |
| 15.4 | Desviación típica | 41 |
| 15.4.1 | Cuasidesviación típica | 41 |
| 15.4.2 | Caracterización de una distribución estadística | 42 |
| 15.5 | Otras medidas de dispersión | 42 |
| 15.5.1 | Recorrido intercuartílico | 42 |
| 15.5.2 | Coeficiente de variación | 43 |
| 15.6 | Dispersión y transformaciones | 47 |
| 15.6.1 | Cambio de origen | 47 |
| 15.6.2 | Cambio de escala | 49 |
| 15.6.3 | Cambio de origen y escala | 51 |
| 15.6.4 | Variable tipificada | 53 |
| | Bibliografía | 54 |

Tema 1

Fenómenos aleatorios. Conceptos de probabilidad. Propiedades. Independencia de sucesos. Teorema de Bayes.

1.1 Historia de la probabilidad

El azar y la incertidumbre han condicionado desde siempre la vida y el comportamiento de los seres humanos. Las variaciones en el clima y su impacto en el abastecimiento de los alimentos han obligado desde el principio al hombre a esforzarse por reducir la incertidumbre y sus efectos. Los juegos de azar tienen también una larga historia; recordemos, por ejemplo, que inmediatamente después de crucificar a Cristo, los soldados se jugaron su túnica a los dados.

Se admite que la correspondencia, a mediados del siglo XVII, entre Antoine Gombaud, caballero de Méré (1607-1684) y Blaise Pascal (1623-1662), sobre el reparto del dinero apostado al principio de una partida que es repentinamente interrumpida, marca el origen del estudio de lo que hoy conocemos como teoría de la probabilidad. A su vez, Pascal introdujo a otro matemático Pierre Fermat (1601-1665) en el reto de resolver el problema que verosímilmente estaba relacionado con la ilegalidad de las partidas en aquella época. No obstante, algunas probabilidades numéricas para ciertas combinaciones de resultados de dados ya habían sido calculadas por Girolamo Cardano (1501-1576) y por Galileo Galilei (1564-1642). La teoría de probabilidad ha sido constantemente desarrollada desde el siglo XVII y ampliamente aplicada en diversos campos de estudio. Hoy, la teoría de la probabilidad es una herramienta importante en la mayoría de las áreas de ingeniería, ciencias y administración.

1.2 Interpretaciones de la probabilidad

Además de muchas aplicaciones formales de la teoría, el concepto de probabilidad aparece en nuestra vida y en nuestras conversaciones cotidianas. Pero a pesar de que dicho concepto es una parte tan común y natural de nuestra experiencia, no existe una única interpretación científica del término probabilidad aceptada por todos los estadísticos, filósofos y demás autoridades científicas. A través de los años, cada interpretación propuesta por unos expertos ha sido criticada por otros. De hecho, el verdadero significado de la probabilidad es todavía un tema muy conflictivo y surge en muchas discusiones filosóficas actuales sobre los fundamentos de la estadística. Describiremos aquí tres interpretaciones diferentes de la probabilidad antes de introducir más adelante, la definición axiomática, la más comúnmente aceptada.

1.2.1 La interpretación frecuentista de la probabilidad

En muchos problemas, la probabilidad de obtener algún resultado específico de un proceso puede ser interpretada en el sentido de la frecuencia relativa con la que se obtendría ese resultado, si el proceso se repitiera un número grande de veces en condiciones similares. Por ejemplo, la probabilidad de obtener una cara cuando se lanza una moneda equilibrada es considerada $\frac{1}{2}$,

debido a que la frecuencia relativa de caras debería ser aproximadamente $\frac{1}{2}$, cuando la moneda es lanzada un número grande de veces en condiciones similares. En otras palabras, se supone que la proporción de lanzamientos en los que se obtiene una cara sería aproximadamente $\frac{1}{2}$.

Está claro que las condiciones mencionadas en este ejemplo son muy vagas para servir como base de una definición científica de probabilidad. En primer lugar, se menciona un “número grande” de lanzamientos de la moneda, pero no hay una indicación clara del número específico que podría considerarse suficientemente grande. En segundo lugar, se afirma que los lanzamientos deberían tener lugar “en condiciones similares”, pero estas condiciones no se describen con precisión. Las condiciones en las cuales se lanza la moneda no pueden ser completamente idénticas para todos los lanzamientos porque entonces los resultados serían todos iguales y se obtendrían sólo caras o sólo cruces. Consecuentemente, los lanzamientos no deben ser completamente controlados sino que deben tener algunas características “aleatorias”.

Se afirma, además, que la frecuencia relativa de caras sería “aproximadamente $\frac{1}{2}$ ”, pero no se especifica un límite para la variación posible respecto del valor $\frac{1}{2}$. Si una moneda fuera lanzada 1,000,000 de veces, no esperaríamos obtener exactamente 500,000 caras. En realidad, nos sorprenderíamos mucho si obtuviéramos exactamente 500,000. Por otro lado, tampoco esperaríamos que el número de caras distara mucho de 500,000. Sería deseable poder hacer una afirmación precisa de las verosimilitudes de los diferentes números posibles de caras, pero estas verosimilitudes dependerían necesariamente del mismo concepto de probabilidad que estamos intentando definir.

Otro inconveniente de la interpretación frecuentista de la probabilidad es que sólo puede utilizarse para un problema en el que pueda haber, al menos en principio, un número grande de repeticiones similares de cierto proceso. Muchos problemas importantes no son de este tipo. Por ejemplo, la interpretación frecuentista de la probabilidad de que una determinada central nuclear falle.

1.2.2 La interpretación clásica de la probabilidad

La interpretación clásica de la probabilidad, Laplace (1749-1827), está basada en el concepto de resultados igualmente verosímiles. Por ejemplo, cuando se lanza una moneda existen dos resultados posibles: cara o cruz. Si se puede suponer que la ocurrencia de estos resultados es igualmente verosímil, entonces deben tener la misma probabilidad. Puesto que la suma de las probabilidades debe ser 1, tanto la probabilidad de una cara como la probabilidad de una cruz debe ser $\frac{1}{2}$. Generalizando, si el resultado de algún proceso debe ser uno de n resultados diferentes, y si estos n resultados son igualmente verosímiles, entonces la probabilidad de cada resultado es $\frac{1}{n}$.

Dos dificultades básicas aparecen cuando se intenta desarrollar una definición formal de probabilidad desde la interpretación clásica. En primer lugar, el concepto de resultados igualmente verosímiles se basa en esencia en el concepto de probabilidad que estamos tratando de definir. Afirmar que dos resultados posibles son igualmente verosímiles es lo mismo que afirmar que

los resultados tienen la misma probabilidad. En segundo lugar, no se proporciona un método sistemático para asignar probabilidades a resultados que no sean igualmente verosímiles. Cuando se lanza una moneda o se arroja un dado equilibrado o se escoge una carta de una baraja bien mezclada, los diferentes resultados posibles pueden, en general, ser considerados igualmente verosímiles debida a la naturaleza del proceso. Sin embargo, cuando el problema es predecir si una persona se casará o si un proyecto de investigación tendrá éxito, los resultados posibles no suelen considerarse igualmente verosímiles, y es necesario un método diferente para asignar probabilidades a estos resultados.

1.2.3 La interpretación subjetiva de la probabilidad

De acuerdo con ella la probabilidad que una persona asigna a cada uno de los posibles resultados de un experimento viene dada por su propio juicio sobre la verosimilitud de cada resultado. Si los juicios de la persona acerca de estas verosimilitudes relativas a diversas combinaciones de resultados satisfacen ciertas condiciones de consistencia, puede demostrarse que las probabilidades subjetivas de los diferentes sucesos quedan determinadas de forma única. Las dificultades asociadas con esta perspectiva son dos: parece humanamente difícil asumir que los juicios de una persona sobre un número infinito de sucesos estén libres de contradicciones internas; por otro lado, no se dan claves para que dos o más personas que trabajan juntas tengan juicios consistentes entre ellas.

1.3 Experimentos y sucesos

La teoría de la probabilidad tiene que ver con los diversos resultados posibles que podrían obtenerse y los posibles sucesos que podrían ocurrir cuando se lleva a cabo un experimento aleatorio. El término “experimento” se utiliza en la teoría de la probabilidad para describir cualquier proceso cuyos resultados no se conocen de antemano con certeza. A continuación se presentan algunos ejemplos de experimentos:

- i) En un experimento en el cual una moneda ha de ser lanzada 10 veces, el experimentador podría estar interesado en determinar la probabilidad de obtener al menos 4 caras.
- ii) En un experimento en el cual va a seleccionarse una muestra de 1.000 transistores de un gran cargamento de artículos similares y en el que se inspeccionará cada artículo seleccionado, una persona podría estar interesada en determinar la probabilidad de que no más de uno de los transistores seleccionados sea defectuoso.
- iii) A partir de información de la vida de Einstein, alguien podría desear establecer la probabilidad de que naciera en el año 1915.

1.4 Teoría de conjuntos

1.4.1 Espacio muestral

El conjunto de todos los posibles resultados de un experimento se llama espacio muestral del experimento. El lenguaje y los conceptos de la teoría de conjuntos proporcionan un contexto natural para el desarrollo de la teoría de probabilidad. A continuación se revisan ideas básicas de la teoría de conjuntos.

1.4.2 Relaciones de la teoría de conjuntos

Sea Ω el espacio muestral de un experimento aleatorio. Se dice que cualquier resultado posible ω del experimento es un miembro del espacio Ω o que pertenece al espacio Ω . La afirmación de que ω es un miembro de Ω se denota simbólicamente por $\omega \in \Omega$.

Cuando se ha llevado a cabo un experimento y se dice que ha ocurrido un suceso, lo que se afirma es que el resultado del experimento satisfizo ciertas condiciones que especifican ese suceso. En otras palabras, algunos resultados del espacio Ω indican que el suceso ocurrió, y los restantes resultados de Ω indican que el suceso no ocurrió. De acuerdo con esta interpretación, cualquier suceso puede ser considerado como un subconjunto de posibles resultados del espacio Ω .

Ejemplo 1.1: Se lanza un dado de seis caras. El espacio muestral contiene los seis números 1, 2, 3, 4, 5 y 6. Simbólicamente, se escribe

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

El suceso S_1 de obtener un número par está definido por el subconjunto

$$S_1 = \{2, 4, 6\}.$$

El suceso S_2 de obtener un número mayor que 2 está definido por el subconjunto

$$S_2 = \{3, 4, 5, 6\}.$$

Se dice que el suceso S_1 está contenido en otro suceso S_2 si cada resultado que pertenece al subconjunto que define el suceso S_1 pertenece al subconjunto que define el suceso S_2 . Esta relación entre dos sucesos se expresa simbólicamente mediante $S_1 \subseteq S_2$. La relación $S_1 \subseteq S_2$ se expresa diciendo que S_1 es un subconjunto de S_2 . Equivalentemente, si $S_1 \subseteq S_2$, puede decirse que S_2 contiene a S_1 y escribir $S_2 \supseteq S_1$.

Ejemplo 1.2: Volvamos al ejemplo anterior, donde se lanzaba un dado un dado de 6 caras. Supongamos que S_1 es el suceso de obtener un número par y S_3 es el suceso de obtener un número mayor que 1. Puesto que $S_1 = \{2, 4, 6\}$ y $S_3 = \{2, 3, 4, 5, 6\}$ resulta que $S_1 \subseteq S_3$.

Obsérvese que $S \subseteq \Omega$ para cualquier suceso S .

Si dos sucesos S_1 y S_2 son tales que $S_1 \subseteq S_2$ y $S_2 \subseteq S_1$, resulta que S_1 y S_2 deben contener exactamente los mismos puntos. En otras palabras $S_1 = S_2$.

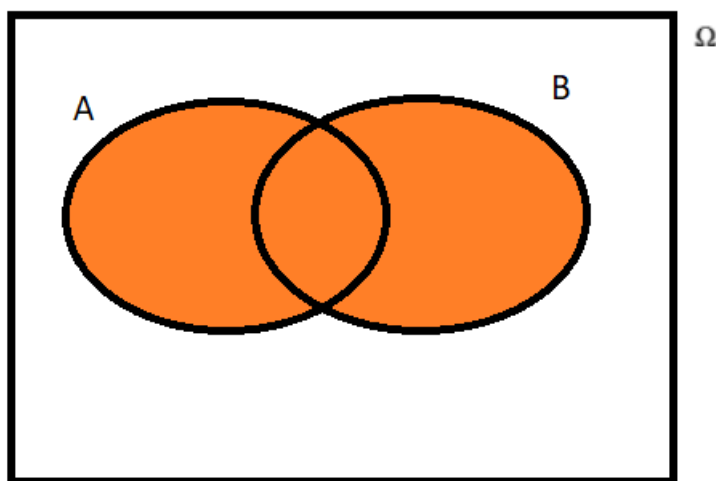
Si S_1 , S_2 y S_3 son tres sucesos tales que $S_1 \subseteq S_2$ y $S_2 \subseteq S_3$, entonces resulta que $S_1 \subseteq S_3$.

1.4.3 El conjunto vacío

Algunos sucesos son imposibles. Por ejemplo, cuando se lanza un dado, es imposible obtener un número negativo. De ahí que el suceso de obtener un número negativo se defina como el subconjunto de Ω que no contiene resultados. Este subconjunto de Ω , se llama conjunto vacío, y se denota por el símbolo \emptyset .

1.4.4 Operaciones de la teoría de conjuntos

Unión. Si A y B son dos sucesos cualesquiera, la unión de A y B se define como el suceso que contiene todos los resultados que pertenecen sólo a A , sólo a B o a ambos, A y B . La notación de la unión de A y B es $A \cup B$.

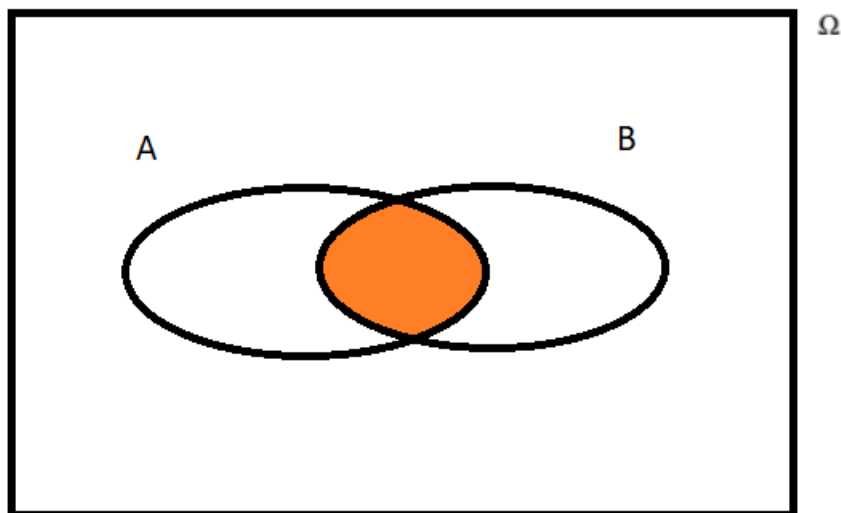


Para cualesquiera sucesos A , B y C , la unión tiene las siguientes propiedades:

- i) $A \cup B = B \cup A$.
- ii) $A \cup A = A$.
- iii) $A \cup \emptyset = A$.
- iv) $A \cup \Omega = \Omega$.
- v) Si $A \subseteq B$, entonces $A \cup B = B$.
- vi) $A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C) = (A \cup C) \cup B$.

Intersección. Para sucesos A y B cualesquiera, la intersección de A y B se define como el suceso que contiene todos los resultados que pertenecen a ambos, A y B . La notación de la intersección

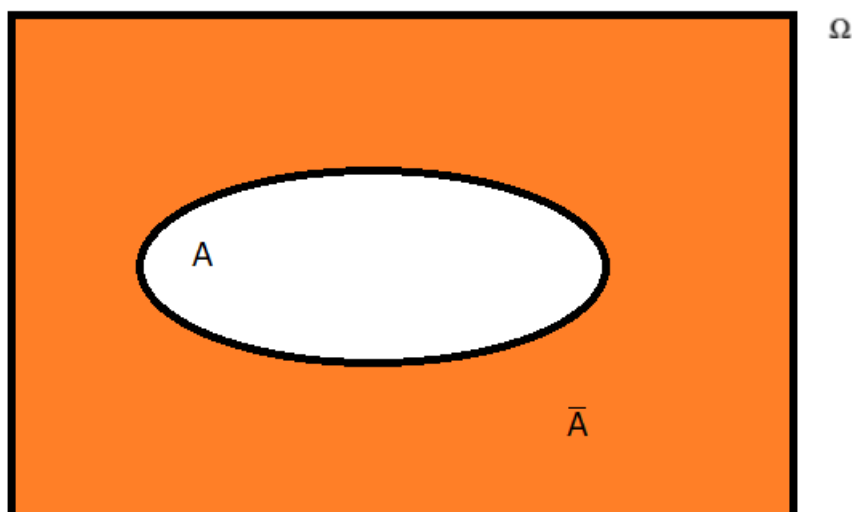
de A y B es $A \cap B$. A menudo es conveniente denotar la intersección de A y B por el símbolo AB en lugar de $A \cap B$, y usaremos indistintamente estos dos tipos de notación.



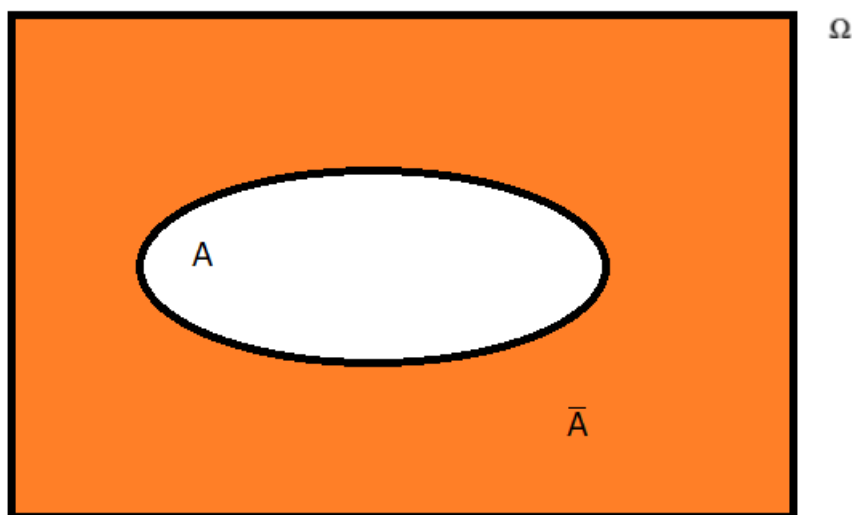
Para cualesquiera sucesos A , B y C , la intersección tiene las siguientes propiedades:

- i) $A \cap B = B \cap A$.
- ii) $A \cap A = A$.
- iii) $A \cap \emptyset = \emptyset$.
- iv) $A \cap \Omega = A$.
- v) Si $A \subseteq B$, entonces $A \cap B = A$.
- vi) $A \cap B \cap C = (A \cap B) \cap C = A \cap (B \cap C) = (A \cap C) \cap B$.

Suceso complementario. El suceso complementario de un suceso A cualquiera se define como el suceso que contiene a todos los resultados del espacio muestral Ω que no pertenecen al suceso A . La notación para el suceso complementario al suceso A es \bar{A} .



Sucesos disjuntos. Se dice que dos sucesos A y B cualesquiera son disjuntos, o mutuamente excluyentes, si A y B no tienen resultados en común. Entonces, A y B son disjuntos si, y sólo si, $A \cap B = \emptyset$.



Leyes de Morgan. Dados dos sucesos A y B en un espacio muestral, se tiene que,

$$i) \overline{A \cup B} = \bar{A} \cap \bar{B}.$$

$$ii) \overline{A \cap B} = \bar{A} \cup \bar{B}.$$

1.5 Definición axiomática de probabilidad

Se presenta a continuación la definición matemática o axiomática de la probabilidad, introducida por el matemático ruso Andréi Kolmogórov en 1933. Para ello necesitamos introducir el concepto de σ -álgebra sobre un conjunto.

Definición 1.1: Dado un conjunto Ω llamaremos σ -álgebra sobre Ω , notada $\sigma(\Omega)$ a cualquier familia de subconjuntos de Ω que verifique:

$$i) \emptyset \in \sigma(\Omega).$$

$$ii) \forall A, B \in \sigma(\Omega) \Rightarrow A \cup B \in \sigma(\Omega).$$

$$iii) \forall A \in \sigma(\Omega) \Rightarrow \bar{A} \in \sigma(\Omega).$$

Definición 1.2 (axiomática de probabilidad): Dado el espacio muestral Ω de un experimento aleatorio y $\sigma(\Omega)$ una σ -álgebra sobre Ω llamaremos probabilidad sobre Ω a cualquier función P definida en $\sigma(\Omega)$ y que asigna a cada $S \in \sigma(\Omega)$ un valor $P(S) \in [0, 1]$ satisfaciendo:

$$i) \text{ La probabilidad del suceso seguro es igual a 1, es decir, } P(\Omega) = 1.$$

ii) Si dos sucesos son disjuntos, la probabilidad de que uno u otro ocurra es la suma de sus probabilidades individuales. De hecho, se supondrá que esta propiedad aditiva de la probabilidad es cierta también para cualquier número finito de sucesos disjuntos y aún para cualquier sucesión infinita de sucesos disjuntos. Es decir, para cualquier sucesión infinita de sucesos disjuntos A_1, A_2, \dots

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

A la terna $(\Omega, \sigma(\Omega), P)$ se le denomina espacio de probabilidad.

1.5.1 Consecuencias de la definición de probabilidad

Se presentarán ahora varias consecuencias importantes de los axiomas anteriores.

Proposición 1.1: Sea $(\Omega, \sigma(\Omega), P)$ un espacio de probabilidad. Entonces,

i) $P(\emptyset) = 0$.

ii) Para cualquier sucesión finita de n sucesos disjuntos A_1, \dots, A_n ,

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

iii) Para cualquier suceso A ,

$$P(\bar{A}) = 1 - P(A).$$

iv) Si $A \subseteq B$, entonces

$$P(A) \leq P(B).$$

v) Para dos sucesos cualesquiera A y B ,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Demostración:

i) Considérese la sucesión infinita de sucesos A_1, A_2, \dots tales que $A_i = \emptyset$ para $i = 1, 2, \dots$, de forma que cada uno de los sucesos es exactamente igual al conjunto vacío. Se trata de una sucesión de sucesos disjuntos, puesto que $\emptyset \cap \emptyset = \emptyset$. Además, $\bigcup_{i=1}^{\infty} A_i = \emptyset$. Por tanto, por el segundo axioma resulta que:

$$P(\emptyset) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} P(\emptyset).$$

Esta ecuación afirma que cuando el número $P(\emptyset)$ se suma repetidamente en una serie infinita, la suma de esa serie es simplemente el número $P(\emptyset)$. El único número real con esta propiedad es $P(\emptyset) = 0$.

ii) Considérese la sucesión finita de sucesos A_1, A_2, \dots, A_n en la cual A_1, \dots, A_n son las n sucesos disjuntos dados y $A_i = \emptyset$ para $i > n$. Entonces los sucesos en esta sucesión finita son disjuntos y $\cap_{i=1}^{\infty} A_i = \cap_{i=1}^n A_i$. Por tanto, utilizando el resultado i),

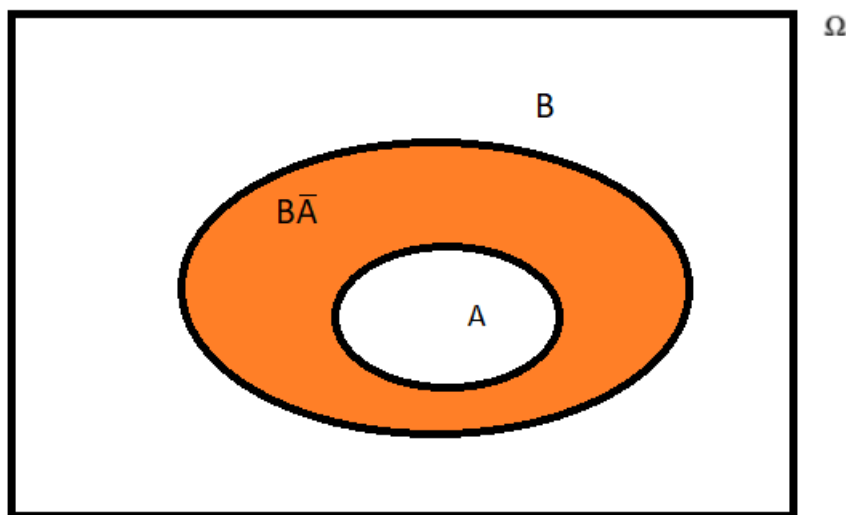
$$P(\cup_{i=1}^{\infty} A_i) = P(\cup_{i=1}^n A_i) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^n P(A_i) + \sum_{i=n+1}^{\infty} P(A_i) = \sum_{i=1}^n P(A_i) + 0 = \sum_{i=1}^n P(A_i).$$

iii) Puesto que A y \bar{A} son sucesos disjuntos y $A \cup \bar{A} = \Omega$, se deduce del ii) que

$$P(\Omega) = P(A) + P(\bar{A}).$$

Puesto que $P(\Omega) = 1$, entonces $P(\bar{A}) = 1 - P(A)$.

iv) Como se ilustra en la siguiente figura, el suceso B puede ser tratado como la unión de los dos sucesos disjuntos A y $B \cap \bar{A}$. Por tanto, $P(B) = P(A) + P(B \cap \bar{A})$. Puesto que $P(B \cap \bar{A}) \geq 0$, entonces $P(B) \geq P(A)$.



v) Como se ilustra en la siguiente figura,

$$A \cup B = (A \cap \bar{B}) \cup (A \cap B) \cup (\bar{A} \cap B).$$

Puesto que los tres sucesos de la parte derecha de esta ecuación son disjuntos, se deduce de ii) que

$$P(A \cup B) = P(A \cap \bar{B}) + P(A \cap B) + P(\bar{A} \cap B).$$

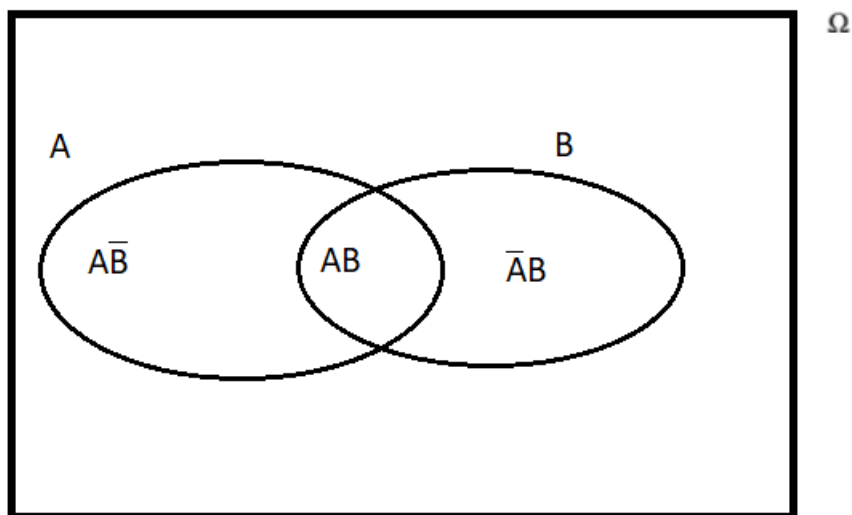
Además, en la siguiente figura también se ve que

$$P(A) = P(A \cap \bar{B}) + P(A \cap B)$$

y

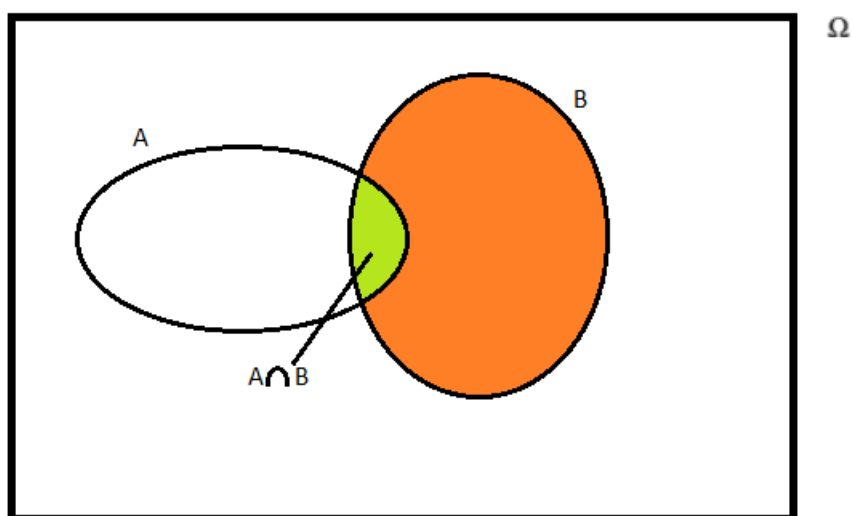
$$P(B) = P(\bar{A} \cap B) + P(A \cap B).$$

El resultado ν) es consecuencia de estas relaciones.



1.6 Probabilidad condicional

Supóngase que se realiza un experimento cuyo espacio muestral de resultados es Ω y también que se han especificado las probabilidades para todos los sucesos de la σ -álgebra. Se estudiará ahora la forma en que cambia la probabilidad de un suceso A cuando se sabe que otro suceso B ha ocurrido. Esta nueva probabilidad de A se denomina probabilidad condicional del suceso A dado que el suceso B ha ocurrido. La notación para esta probabilidad condicional es $P(A|B)$. Si se sabe que un suceso B ha ocurrido, entonces se sabe que el resultado del experimento es uno de los incluidos en B . Por tanto, para evaluar la probabilidad de que A ocurra, se debe considerar el conjunto de los resultados incluidos en B que también implican la ocurrencia de A . Como se representa a continuación, este conjunto es precisamente el conjunto $A \cap B$.



Resulta, por tanto, natural definir la probabilidad condicional $P(A|B)$ como la proporción de la probabilidad total $P(B)$ representada por la probabilidad $P(A \cap B)$. Estas consideraciones conducen a la siguiente definición.

Definición 1.3: Sea $(\Omega, \sigma(\Omega), P)$ un espacio probabilístico y sean A y B son dos sucesos cualesquiera de $\sigma(\Omega)$ con $P(B) > 0$, entonces,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

La probabilidad condicional $P(A|B)$ no esta definida si $P(B) = 0$, dado que no tiene sentido condicionar a algo imposible.

La probabilidad condicional $P(A|B)$ tiene una interpretación sencilla desde el punto de vista de la concepción frecuentista de la probabilidad. En efecto, de acuerdo con esa interpretación, si un proceso experimental se repite un número grande de veces, entonces la proporción de repeticiones en las cuales el suceso B ocurre es aproximadamente $P(B)$ y la proporción de repeticiones en las cuales el suceso A y el suceso B ocurren es aproximadamente $P(A \cap B)$. Por tanto, entre las repeticiones en que el suceso B ocurre, la proporción de repeticiones en que también ocurre el suceso A es aproximadamente igual a

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Ejemplo 1.3: Supóngase que se han lanzado dos dados y que se ha observado que la suma de los dos números ha sido impar. Se determinará la probabilidad de que haya sido menor que 8.

Si se define A como el suceso “la suma es menor que 8” y B como suceso de que “la suma es impar”, entonces $A \cap B$ es el suceso de que “la suma es 3, 5 ó 7”. Las probabilidades respectivas de suma 3, 5 y 7 son: $\frac{2}{36}$, $\frac{4}{36}$ y $\frac{6}{36}$.

A partir de estas probabilidades se calcula $P(A \cap B)$ y $P(B)$ como sigue:

$$P(A \cap B) = \frac{2}{36} + \frac{4}{36} + \frac{6}{36} = \frac{1}{3}$$

y

$$P(B) = \frac{2}{36} + \frac{4}{36} + \frac{6}{36} + \frac{4}{36} + \frac{2}{36} = \frac{18}{36} = \frac{1}{2}.$$

Por tanto,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2}{3}.$$

1.7 Sucesos independientes

Supóngase que dos sucesos A y B ocurren independientemente uno del otro, en el sentido de que la ocurrencia, o no ocurrencia de cualquiera de ellos no tiene relación ni influye sobre la

ocurrencia, o no ocurrencia del otro. Entonces $P(A|B) = P(A)$. Puesto que, por definición la $P(A|B) = \frac{P(A \cap B)}{P(B)}$, para sucesos independientes se tiene,

$$P(A \cap B) = P(A)P(B).$$

Este resultado se puede justificar fácilmente desde el punto de vista de la interpretación frecuencionalista de la probabilidad. Por ejemplo, supóngase que A es el suceso de obtener una cara cuando se lanza una moneda equilibrada y B es el suceso de obtener el número 1 o el número 2 al lanzar un dado equilibrado. Entonces, el suceso A ocurrirá con una frecuencia relativa de $\frac{1}{2}$ al lanzar la moneda repetidamente y el suceso B ocurrirá con una frecuencia relativa de $\frac{1}{3}$ al lanzar el dado repetidamente. Por tanto,

$$P(A) = \frac{1}{2} \text{ y } P(B) = \frac{1}{3}.$$

Considérese ahora un experimento compuesto en el cual la moneda y el dado son lanzados simultáneamente. Si el experimento se realiza repetidamente, entonces la frecuencia relativa del suceso A , en el cual se obtiene una cara, será $\frac{1}{2}$. Puesto que los resultados de la moneda y los resultados del dado no están relacionados, es razonable suponer que entre los dos experimentos en los que el suceso A ocurre, el suceso B en el que se obtiene el número 1 o el número 2 tendrá una frecuencia relativa de $\frac{1}{3}$. De aquí que, en una sucesión de experimentos compuestos, la frecuencia relativa con que ocurren A y B simultáneamente será $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$. Por tanto, en este experimento,

$$P(A \cap B) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = P(A)P(B).$$

Estas relaciones se pueden justificar también mediante la interpretación clásica de la probabilidad.

Existen dos resultados igualmente probables del lanzamiento de la moneda y seis resultados igualmente probables del lanzamiento de un dado. Puesto que el resultado de la moneda y el resultado del dado no están relacionados, es natural suponer también que los $6 \times 2 = 12$ resultados posibles del experimento compuesto en el cual se lanzan la moneda y el dado son igualmente probables. En este experimento compuesto será cierto de nuevo que,

$$P(A \cap B) = \frac{1}{6} = P(A)P(B).$$

Ejemplo 1.4: Supóngase que dos máquinas, 1 y 2, de una fábrica funcionan independientemente una de la otra. Sea A el suceso de que la máquina 1 se estropee durante un periodo determinado de 8 horas y sea B el suceso de que la máquina 2 se estropee durante el mismo periodo y supóngase que $P(A) = \frac{1}{3}$ y $P(B) = \frac{1}{4}$. Se determinará la probabilidad de que al menos una de las máquinas se estropee durante ese periodo.

La probabilidad $P(A \cap B)$ de que ambas máquinas se estropeen durante ese periodo es,

$$P(A \cap B) = P(A)P(B) = \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}.$$

Por tanto, la probabilidad $P(A \cup B)$ de que al menos una de las máquinas se estropee durante ese periodo es,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{3} + \frac{1}{4} - \frac{1}{12} = \frac{1}{2}.$$

En el ejemplo siguiente se muestra que dos sucesos A y B , que están relacionados físicamente pueden satisfacer, sin embargo, la definición de independencia.

Ejemplo 1.5: Supóngase que se lanza un dado equilibrado. Sea A el suceso de obtener un número par y sea B el suceso de obtener uno de los números 1, 2, 3 ó 4. Demostraremos que los sucesos A y B son independientes.

En este ejemplo, $P(A) = \frac{1}{2}$ y $P(B) = \frac{1}{3}$. Además, puesto que $A \cap B$ es el suceso de obtener el número 2 o el número 4, $P(A \cap B) = \frac{1}{3}$. De aquí que $P(A \cap B) = P(A)P(B)$. Resulta pues que los sucesos A y B son independientes, aunque la ocurrencia de cada suceso depende del mismo lanzamiento del dado.

La independencia de los sucesos A y B del ejemplo 1.4 también se puede interpretar como sigue: Supóngase que una persona debe apostar sobre si el número obtenido al lanzar el dado es par o impar, es decir, sobre si ocurrirá el suceso A . Puesto que tres de los posibles resultados del lanzamiento son pares y los otros tres son impares, en general, esa persona no tendrá preferencia entre apostar por un número par o apostar por un número impar. Supóngase también que después de haber lanzado el dado, pero antes de que la persona se entere del resultado y antes de que se decida apostar sobre un resultado par o impar, es informada de que el resultado ha sido uno de los números 1, 2, 3 ó 4, es decir, que ha ocurrido el suceso B . Esa persona sabe ahora que el resultado ha sido 1, 2, 3 ó 4. Sin embargo, puesto que dos de estos número son pares y los otros dos son impares, aún no tendrá preferencia entre apostar por un número par o apostar por un número impar. En otras palabras, la información de que ha ocurrido el suceso B no sirve de ayuda a la persona que está tratando de decidir si ha ocurrido el suceso A .

Proposición 1.2: Si dos sucesos A y B son independientes, entonces los sucesos A y \bar{B} también son independientes.

Demostración: Siempre es cierto que,

$$P(A \cap \bar{B}) = P(A) - P(A \cap B).$$

Además, puesto que A y B son sucesos independientes,

$$P(A \cap B) = P(A)P(B).$$

Resulta que,

$$P(A \cap \bar{B}) = P(A) - P(A)P(B) = P(A)[1 - P(B)] = P(A)P(\bar{B}).$$

Por tanto, los sucesos A y \bar{B} son independientes.

La demostración del resultado análogo para los sucesos \bar{A} y B es similar.

Proposición 1.3: Si dos sucesos A y B son independientes, entonces los sucesos \bar{A} y \bar{B} también son independientes.

Demostración: $P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B)$.

La primera igualdad se debe a la ley de Morgan. Entonces,

$$P(\bar{A} \cap \bar{B}) = 1 - [P(A) + P(B) - P(A \cap B)] = 1 - P(A) - P(B) + P(A)P(B),$$

la última igualdad por la independencia de los sucesos A y B .

Finalmente $P(\bar{A} \cap \bar{B}) = (1 - P(A))(1 - P(B)) = P(\bar{A})P(\bar{B})$.

Independencia de varios sucesos

La noción de independencia de dos sucesos se puede extender a cualquier número de ellos. Si k sucesos A_1, \dots, A_k son independientes, entonces es natural suponer que la probabilidad $P(A_1 \cap \dots \cap A_k)$ de que los k sucesos ocurran es el producto $P(A_1) \cdots P(A_k)$. Además, puesto que los sucesos A_1, \dots, A_k no están relacionados, esta regla de producto debería cumplirse no sólo para la intersección de k sucesos, sino también para la intersección de dos cualesquiera de ellos, o de tres cualesquiera de ellos o de cualquier otro número de ellos. Estas consideraciones conducen a la siguiente definición.

Definición 1.4: Los k sucesos A_1, \dots, A_k son independientes si para cada subconjunto A_{i_1}, \dots, A_{i_j} de estos sucesos ($j = 2, 3, \dots, k$),

$$P(A_{i_1} \cap \dots \cap A_{i_j}) = P(A_{i_1}) \cdots P(A_{i_j}).$$

En particular, para que tres sucesos A , B y C , sean independientes, deben satisfacerse las cuatro relaciones siguientes:

$$P(A \cap B) = P(A)P(B), P(A \cap C) = P(A)P(C), P(B \cap C) = P(B)P(C), \quad (1.1)$$

y

$$P(A \cap B \cap C) = P(A)P(B)P(C) \quad (1.2)$$

Ejemplo 1.6: Considérese un experimento en el cual el espacio muestral Ω contiene cuatro resultados $\{\omega_1, \omega_2, \omega_3, \omega_4\}$ y supóngase que la probabilidad de cada resultado es $\frac{1}{4}$. Considérense los sucesos A , B y C definidos como sigue:

$$A = \{\omega_1, \omega_2\},$$

$$B = \{\omega_1, \omega_3\},$$

$$C = \{\omega_1, \omega_4\}.$$

Entonces $A \cap B = A \cap C = B \cap C = A \cap B \cap C = \{\omega_1\}$. Por tanto,

$$P(A) = P(B) = P(C) = \frac{1}{2},$$

y

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = P(A \cap B \cap C) = P(\{\omega_1\}) = \frac{1}{4}.$$

Resulta que aunque cada una de las tres relaciones 1.1 se verifica, no se cumple la ecuación 1.2. Diremos, en este caso, que los sucesos A , B y C son independientes por pares, pero los tres no son independientes.

A continuación se presentan algunos ejemplos que ilustran la potencia y el alcance del concepto de independencia en la solución de problemas de probabilidad.

Ejemplo 1.7: Supóngase que una máquina produce un artículo defectuosos con probabilidad $p \in [0, 1]$ y produce un artículo no defectuoso con probabilidad $q = 1 - p$. Supóngase además que se seleccionan aleatoriamente para su control seis de los artículos producidos por la máquina y que los resultados de control son independientes para estos seis artículos. Se determinará la probabilidad de que exactamente dos de los seis artículos sean defectuosos.

Se puede suponer que el espacio muestral Ω contiene todas las configuraciones posibles de los resultados de control de seis artículos, para uno de los cuales puede ser defectuoso o no defectuoso. Para $j = 1, \dots, 6$, se definirá D_j como el suceso de que el artículo j -ésimo en la muestra sea defectuoso y N_j como el suceso de que este artículo no sea defectuoso. Puesto que los resultados del control de seis artículos distintos son independientes, la probabilidad de obtener cualquier sucesión concreta de artículos defectuosos o no defectuosos será simplemente el producto de las probabilidades individuales de los resultados del control de cada artículo. Por ejemplo,

$$\begin{aligned} P(N_1, D_2, N_3, N_4, D_5, N_6) &= P(N_1)P(D_2)P(N_3)P(N_4)P(D_5)P(N_6) \\ &= qpqqpq = p^2q^4. \end{aligned}$$

Es fácil ver que la probabilidad de cualquier otra sucesión concreta de Ω formada por dos artículos defectuosos y cuatro no defectuosos será también p^2q^4 . Por tanto, la probabilidad de que en una muestra de seis artículos haya exactamente dos defectuosos se puede obtener multiplicando la probabilidad p^2q^4 de cualquier sucesión concreta que contenga dos defectuosos por el número posible de tales sucesiones. Puesto que hay $\binom{6}{2}$ configuraciones distintas de dos artículos defectuosos y cuatro no defectuosos, la probabilidad de obtener exactamente dos defectuosos es $\binom{6}{2} p^2q^4$.

Ejemplo 1.8: Partiendo de la hipótesis del ejemplo anterior, se determinará ahora la probabilidad de que al menos uno de los seis artículos de la muestra sea defectuoso.

Puesto que los resultados del control de los diferentes artículos son independientes, la probabilidad de que los seis artículos sean no defectuosos es q^6 . Por tanto, la probabilidad de que al menos un artículo sea defectuoso es $1 - q^6$.

Ejemplo 1.9: Supóngase que se lanza una moneda equilibrada hasta que aparece una cara por primera vez, y supóngase que los resultados de los lanzamientos son independientes. Se determinará la probabilidad p_n de que se necesiten exactamente n lanzamientos.

La probabilidad deseada es igual a la probabilidad de obtener $n - 1$ cruces seguidas y luego obtener una cara en el siguiente lanzamiento. Como los resultados de los lanzamientos son independientes, la probabilidad de esta sucesión concreta de n resultados es $p_n = (\frac{1}{2})^n$.

La probabilidad de obtener una cara antes o después (es decir, de no obtener siempre cruces) es

$$\sum_{n=1}^{\infty} p_n = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1.$$

Puesto que la suma de probabilidades p_n es 1, resulta que la probabilidad de obtener una sucesión infinita de cruces sin obtener nunca una cara debe ser 0.

Ejemplo 1.10: Considérese de nuevo una máquina que produce un artículo defectuoso con probabilidad p y uno no defectuoso con probabilidad $q = 1 - p$. Supóngase que el control se realiza seleccionando artículos al azar y de uno en uno hasta obtener exactamente cinco artículos defectuosos. Se determinará la probabilidad P_n de que deban ser seleccionados exactamente n artículos ($n \geq 5$) para obtener los 5 defectuosos.

El quinto artículo defectuoso será el n -ésimo artículo controlado si, y sólo si, hay exactamente cuatro defectuosos entre los primeros $n - 1$ artículos y el n -ésimo artículo es defectuoso. Siguiendo un razonamiento análogo al utilizado en el ejemplo anterior de las máquinas, se puede demostrar que la probabilidad de obtener exactamente 4 artículos defectuosos y $n - 5$ no defectuosos entre los primeros $n - 1$ artículos es $\binom{n-1}{4} p^4 q^{n-5}$. La probabilidad de que el n -ésimo artículo sea defectuoso es p . Puesto que el primer suceso se refiere únicamente a los resultados del control de los primeros $n - 1$ artículos y el segundo suceso se refiere sólo al resultado del control n -ésimo artículo, estos dos sucesos son independientes. Por tanto, la probabilidad de que ambos sucesos ocurran es igual al producto de sus probabilidades. Resulta, por tanto, que

$$P_n = \binom{n-1}{4} p^5 q^{n-5}.$$

1.7.1 Ley multiplicativa para probabilidades condicionales

En un experimento que involucra dos sucesos A y B que no son independientes, a menudo, es conveniente calcular la probabilidad $P(A \cap B)$ de que ambos sucesos ocurran utilizando una de las ecuaciones siguientes:

$$P(A \cap B) = P(B)P(A|B)$$

ó

$$P(A \cap B) = P(A)P(B|A).$$

Estas expresiones se pueden generalizar al caso de un número arbitrario de sucesos A_1, A_2, \dots, A_n . Entonces, $P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$.

Ejemplo 1.11: En una urna hay 8 bolas, tres de ellas blancas, cuatro azules y una roja. Si se extraen sin remplazamiento sucesivamente tres bolas, la probabilidad de que la primera sea azul, la segunda sea roja y la tercera blanca es

$$P(A_1 \cap R_2 \cap B_3) = P(A_1)P(R_2|A_1)P(B_3|A_1 \cap R_2) = \frac{4}{8} \cdot \frac{1}{7} \cdot \frac{3}{6} = \frac{1}{28}.$$

1.8 Teorema de Bayes

El teorema de Bayes, en la teoría de la probabilidad, es una fórmula planteada por el matemático y clérigo inglés Thomas Bayes (1702-1761) y publicada póstumamente en 1763, que expresa la probabilidad condicional de un evento aleatorio A dado B en términos de la de probabilidad condicional del evento B dado A y la de probabilidad de A .

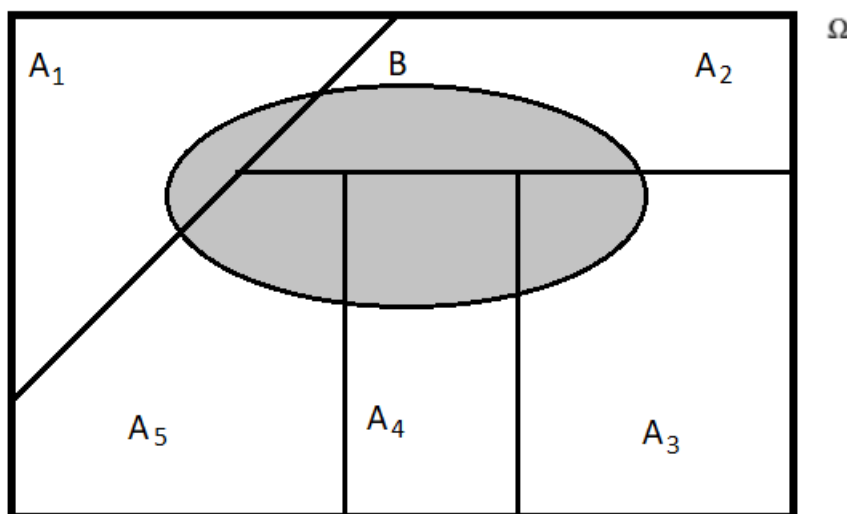
En términos más generales y menos matemáticos, el teorema de Bayes es de enorme relevancia puesto que vincula la probabilidad de A dado B con la probabilidad de B dado A . Es decir, por ejemplo, sabiendo la probabilidad de tener dolor de cabeza dado que se tiene gripe, se podría conocer (si se tiene algún dato más), la probabilidad de tener gripe si se tiene un dolor de cabeza. Muestra este sencillo ejemplo la alta relevancia del teorema en cuestión para la ciencia en todas sus ramas, puesto que tiene vinculación íntima con la comprensión de la probabilidad de aspectos causales dados los efectos observados.

En el magnífico libro *La teoría que nunca murió* de Sharon Bertsch McGrayne (2012) se puede encontrar tanto la historia de la regla de Bayes como un elenco inagotable de aplicaciones en casos reales.

1.8.1 Probabilidad y particiones

Sea Ω el espacio muestral de un experimento y considérense k sucesos A_1, \dots, A_k de Ω de forma que A_1, \dots, A_k sean disjuntos (o mutuamente excluyentes) y tal que $\cup_{i=1}^k A_i = \Omega$, es decir, sean exhaustivos. Se dice que estos sucesos constituyen una partición de Ω .

Si los k sucesos A_1, \dots, A_k constituyen una partición de Ω y si B es cualquier otro suceso en Ω , entonces los sucesos $A_1 \cap B, \dots, A_k \cap B$ constituyen una partición de B , como se muestra a continuación.



Por tanto, se puede escribir

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \dots \cup (A_k \cap B).$$

Además, puesto que los k sucesos del lado derecho de esta ecuación son disjuntos,

$$P(B) = \sum_{j=1}^k P(A_j \cap B) = \sum_{j=1}^k P(A_j)P(B|A_j).$$

En resumen, se ha obtenido el resultado siguiente.

Teorema de la probabilidad total: Supóngase que los sucesos A_1, \dots, A_k forman una partición del espacio Ω y que $P(A_j) > 0$ para $j = 1, \dots, k$. Entonces para cualquier suceso B de Ω ,

$$P(B) = \sum_{j=1}^k P(A_j)P(B|A_j).$$

Ejemplo 1.12: Dos cajas contienen cerrojos grandes y cerrojos pequeños. Supongamos que una caja contienen 60 cerrojos grandes y 40 cerrojos pequeños y que la otra contiene 10 grandes y 20 pequeños. Supóngase también que se selecciona una caja al azar y que se extrae un cerrojo aleatoriamente de la misma. Se determinará la probabilidad de que este cerrojo sea grande.

Sea A_1 el suceso de que se seleccione la primera caja; sea A_2 el suceso de que se seleccione la segunda caja y sea B el suceso de que se seleccione un cerrojo grande. Entonces

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2).$$

Puesto que se selecciona una caja al azar, se sabe que $P(A_1) = P(A_2) = \frac{1}{2}$. Además la probabilidad de seleccionar un cerrojo grande de la primera caja viene dado por $P(B|A_1) = \frac{60}{100} = \frac{3}{5}$ y la

probabilidad de seleccionar un cerrojo grande de la segunda caja es $P(B|A_2) = \frac{10}{30} = \frac{1}{3}$. Por tanto,

$$P(B) = \frac{1}{2} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{1}{3} = \frac{7}{15}.$$

Ejemplo 1.13: Supóngase que en un juego una persona puede obtener como puntuación uno de los 50 números $1, 2, \dots, 50$ y que todos estos números son puntuaciones igualmente verosímiles. La primera vez que juega, su puntuación es i . Entonces continua jugando hasta que obtiene otra puntuación j tal que $j \geq i$. Se puede suponer que todos los juegos son independientes. Se determinará la probabilidad de que j sea 50. Notemos B al suceso la puntuación es 50 y A_i para $i = 1, \dots, 50$ el suceso la puntuación es igual i . Entonces, dado cualquier valor i , el valor j puede ser cualquiera de los números $i, i+1, \dots, 50$. Puesto que estos $(51-i)$ valores posibles de j son igualmente verosímiles, resulta que

$$P(B|A_i) = \frac{1}{51-i}.$$

Además, como la probabilidad de cada uno de los 50 valores posibles de i es $\frac{1}{50}$, resulta que

$$P(B) = \sum_{i=1}^{50} \frac{1}{50} \cdot \frac{1}{51-i} = \frac{1}{50} \left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{50}\right) = 0,09.$$

Ahora se puede establecer el siguiente resultado que es conocido como el teorema de Bayes.

Teorema: Supóngase que los sucesos A_1, \dots, A_k constituyen una partición del espacio Ω tal que $P(A_j) > 0$ para $j = 1, \dots, k$ y sea B cualquier suceso tal que $P(B) > 0$. Entonces, para $i = 1, \dots, k$,

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^k P(A_j)P(B|A_j)}. \quad (1.3)$$

Demostración: Por la definición de probabilidad condicional,

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}.$$

El numerador de la parte derecha de la ecuación 1.3 es igual a $P(A_i \cap B)$ y el denominador es $P(B)$ como consecuencia del teorema de la probabilidad total.

Ejemplo 1.13: Para la fabricación de un gran lote de artículos similares se utilizaron tres máquinas M_1, M_2 y M_3 . Supóngase que el 20 % de los artículos fueron fabricados por la máquina M_1 , el 30 % por la máquina M_2 y el 50 % por la máquina M_3 . Supóngase además que el 1 % de los artículos fabricados por la máquina M_1 son defectuosos, que el 2 % de los artículos fabricados por la máquina M_2 son defectuosos y que el 3 % de los artículos fabricados por la máquina M_3 son defectuosos. Por último, supóngase que se selecciona al azar uno de los artículos del lote y que resulta ser defectuoso. Se determinará la probabilidad de que este artículo haya sido fabricado por la máquina M_2 .

Sea A_i , el suceso de que el artículo seleccionado haya sido fabricado por la máquina M_i ($i = 1, 2, 3$) y sea B el suceso de que el artículo seleccionado sea defectuoso. Hay que calcular la probabilidad condicional $P(A_2|B)$.

La probabilidad $P(A_i)$ de que un artículo seleccionado al azar haya sido fabricado por la máquina M_i es, para $i = 1, 2, 3$:

$$P(A_1) = 0,2,$$

$$P(A_2) = 0,3$$

y

$$P(A_3) = 0,5.$$

Además la probabilidad $P(B|A_i)$ de que un artículo producido por la máquina M_i sea defectuoso es:

$$P(B|A_1) = 0,01,$$

$$P(B|A_2) = 0,02$$

y

$$P(B|A_3) = 0,03.$$

Del teorema de Bayes resulta que

$$\begin{aligned} P(A_2|B) &= \frac{P(A_2)P(B|A_2)}{\sum_{j=1}^k P(A_j)P(B|A_j)} \\ &= \frac{(0,3)(0,02)}{(0,2)(0,01) + (0,3)(0,02) + (0,5)(0,03)} = 0,26. \end{aligned}$$

Bibliografía

1. A.A. Borovkov (2013). *Probability Theory*. Springer.
2. A. Klenke (2014). *Probability Theory. A Comprehensive Course*. Springer.
3. M.H. DeGroot (1988). *Probability and Statistics*. Wilmington, Delaware, E.U.A. Addison-Wesley Iberoamericana, S.A.
4. V.K. Rohatgi and A.K. Md. Ehsanes Saleh (2001). *An Introduction to Probability and Statistics*. John Wiley and Sons.

Tema 2

Variables aleatorias. Variables discretas. Función de probabilidad. Variables continuas. Función de densidad. Propiedades.

2.1 Variables aleatorias.

En probabilidad y estadística, una variable aleatoria es una función que asigna valores reales a los resultados de un experimento aleatorio.

Empecemos recordando la definición de σ -álgebra sobre un conjunto y la de espacio probabilístico.

Definición 2.1: Dado un conjunto Ω llamaremos σ -álgebra sobre Ω , notada $\sigma(\Omega)$ a cualquier familia de subconjuntos de Ω que verifique:

- i) $\emptyset \in \sigma(\Omega)$.
- ii) $\forall A, B \in \sigma(\Omega) \Rightarrow A \cup B \in \sigma(\Omega)$.
- iii) $\forall A \in \sigma(\Omega) \Rightarrow \bar{A} \in \sigma(\Omega)$.

Cuando Ω es un espacio muestral finito, la σ -álgebra habitual es $2^\Omega = \{S | S \subseteq \Omega\}$, es decir, la familia de todos los subconjuntos del espacio muestral.

Definición 2.2: La σ -álgebra de Borel en \mathbb{R} es la familia de subconjuntos de \mathbb{R} que se pueden generar a partir de los conjuntos de la forma $(-\infty, x]$, cuando $x \in \mathbb{R}$. Es decir, la σ -álgebra de Borel en \mathbb{R} contendrá a todos los conjuntos que se puedan obtener como mínimos, máximos, intersecciones ó complementarios de conjuntos de la forma $(-\infty, x]$, $x \in \mathbb{R}$.



Definición 2.3: Un espacio probabilístico asociado a un experimento aleatorio es una terna $(\Omega, \sigma(\Omega), P)$, donde:

- i) Ω es el espacio muestral de los resultados de un experimento aleatorio.
- ii) $\sigma(\Omega)$ es una σ -álgebra sobre Ω .
- iii) P es una probabilidad, es decir:

$$P : \sigma(\Omega) \longrightarrow [0, 1]$$

$$A \longrightarrow P(A),$$

con $P(\emptyset) = 0$ y $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, cuando $\{A_i\}_{i=1}^{\infty}$ son conjuntos disjuntos.

Definición 2.4: Dado un espacio probabilístico $(\Omega, \sigma(\Omega), P)$ llamaremos variable aleatoria X a cualquier función:

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ \omega \in \Omega &\longrightarrow X(\omega), \end{aligned}$$

tal que el conjunto $\forall x \in \mathbb{R}, \{\omega | X(\omega) \leq x\} = \{X \leq x\} \in \sigma(\Omega)$.

Ejemplo 2.1: Consideramos el experimento aleatorio consistente en lanzar dos dados equilibrados de forma independiente. Sea X la variable aleatoria suma de los puntos obtenidos.

$$\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (6, 6)\}, |\Omega| = 6^2 = 36$$

$$P : 2^\Omega \longrightarrow [0, 1]$$

$$A \longrightarrow P(A) = \frac{\text{casos favorables de } A}{\text{casos posibles}} = \frac{|A|}{36}, \text{ (Regla de Laplace).}$$

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ (i, j) &\longrightarrow (i + j), i, j \in \{1, 2, 3, 4, 5, 6\} \end{aligned}$$

Por ejemplo, si $x = 4$, $\{\omega | X(\omega) \leq 4\} = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)\} \in \sigma(\Omega)$.

Como $\sigma(\Omega)$ es 2^Ω entonces en $\sigma(\Omega)$ estarán todos los subconjuntos de Ω y $\forall x, \{X \leq x\} \in \sigma(\Omega)$, por tanto, X es una variable aleatoria.

2.2 Función de distribución de una variable aleatoria.

Definición 2.5: Dado un espacio probabilístico $(\Omega, \sigma(\Omega), P)$ y una variable aleatoria X , llamaremos función de distribución de X a la función real de variable real

$$\begin{aligned} F : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longrightarrow P(X \leq x) \end{aligned}$$

.

Ejemplo 2.2: En el ejemplo anterior (suma de puntos del lanzamiento de dos dados) tenemos que:

$$F(4) = P(X \leq 4) = P(\{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)\}) = \frac{1}{6}.$$

2.2.1 Propiedades de la función de distribución.

Sea X una variable aleatoria definida en un espacio probabilístico $(\Omega, \sigma(\Omega), P)$ y F su función de distribución. Entonces:

$$i) \forall x \in \mathbb{R}, F(x) \in [0, 1].$$

Demostración: $\forall x \in \mathbb{R}, F(x) = P(X \leq x) \in [0, 1]$, por ser una probabilidad..

$$ii) F \text{ es no decreciente, es decir, si } x \leq x_0, F(x) \leq F(x_0).$$

Demostración: Si $x \leq x_0$, $F(x) = P(\{\omega | X(\omega) \leq x\})$ y $F(x_0) = P(\{\omega | X(\omega) \leq x_0\})$. Ahora bien, como $x \leq x_0$, $\{\omega | X(\omega) \leq x\} \subseteq \{\omega | X(\omega) \leq x_0\}$ por lo que se obtiene que $P(\{\omega | X(\omega) \leq x\}) \leq P(\{\omega | X(\omega) \leq x_0\})$ o equivalentemente $F(x) \leq F(x_0)$.

$$iii) \lim_{x \rightarrow \infty} F(x) = \lim_{x \rightarrow \infty} P(X \leq x) = 1$$

$$iv) \lim_{x \rightarrow -\infty} F(x) = \lim_{x \rightarrow -\infty} P(X \leq x) = P(\emptyset) = 0.$$

Variable aleatoria discreta. Función de probabilidad

Definición 2.6: Diremos que la variable aleatoria X es discreta si toma una cantidad finita o numerable de valores.

Para variables aleatorias discretas se define la llamada función de probabilidad (masa o cuantía) como:

$$P_i = P(X = i), i \in I$$

siendo I el conjunto de valores que tomará X . Dicha función satisface:

$$i) P_i = P(X = i) \geq 0, \forall i \in I.$$

$$ii) \sum_{i \in I} P_i = \sum_{i \in I} P(X = i) = 1.$$

Ambas condiciones se deducen trivialmente de la definición de probabilidad ya que:

$$i) P_i = P(X = i) \geq 0, \text{ dado que es una probabilidad.}$$

$$ii) \sum_{i \in I} P(X = i) = P(\Omega) = 1.$$

Ejemplo 2.3: Se lanzan dos dados. Sea X la suma de puntos obtenidos e Y la diferencia entre los puntos del primer dado y el segundo (con signo). Obtengamos la funciones de probabilidad

de X y de Y .

Si las variables no toman demasiados valores se acostumbra a escribir la función de probabilidad en forma de tabla.

| X | $P(X=i).$ |
|-----|-----------|
| 2 | $1/36$ |
| 3 | $2/36$ |
| 4 | $3/36$ |
| 5 | $4/36$ |
| 6 | $5/36$ |
| 7 | $6/36$ |
| 8 | $5/36$ |
| 9 | $4/36$ |
| 10 | $3/36$ |
| 11 | $2/36$ |
| 12 | $1/36$ |

| Y | $P(Y=j).$ |
|-----|-----------|
| -5 | $1/36$ |
| -4 | $2/36$ |
| -3 | $3/36$ |
| -2 | $4/36$ |
| -1 | $5/36$ |
| 0 | $6/36$ |
| 1 | $5/36$ |
| 2 | $4/36$ |
| 3 | $3/36$ |
| 4 | $2/36$ |
| 5 | $1/36$ |

Las variables X e Y no están igualmente distribuidas, dado que aunque tengan las mismas probabilidades, estas no corresponden a los mismos valores.

Ejemplo 2.4: Se lanzan dos dados. Sea X el resultado del primer dado e Y el máximo de los puntos obtenidos en ambos dados. Obtener la función de probabilidad de cada variable.

| X | $P(X=i).$ |
|-----|-----------|
| 1 | $1/6$ |
| 2 | $1/6$ |
| 3 | $1/6$ |
| 4 | $1/36$ |
| 5 | $1/6$ |
| 6 | $1/6$ |

| Y | P(Y=j). |
|---|---------|
| 1 | 1/36 |
| 2 | 3/36 |
| 3 | 5/36 |
| 4 | 7/36 |
| 5 | 9/36 |
| 6 | 11/36 |

Ejemplo 2.5: Se lanzan dos dados. Sea X el mínimo de los puntos obtenidos en ambos dados. Obtener la función de probabilidad de X .

| X | P(X=i). |
|---|---------|
| 1 | 11/36 |
| 2 | 9/36 |
| 3 | 7/36 |
| 4 | 5/36 |
| 5 | 3/36 |
| 6 | 1/36 |

Variables aleatorias absolutamente continuas. Función de densidad

Definición 2.7:

Diremos que una variable aleatoria X es absolutamente continua cuando su función de distribución $F(x)$ es continua, es decir, cuando $\forall x_0 \in \mathbb{R}$, $\lim_{x \rightarrow x_0} F(x) = F(x_0)$.

En tal caso existe una función $f(x)$, llamada función de densidad, tal que $\forall x \in \mathbb{R}$

$$F(x) = \int_{-\infty}^x f(s)ds \Leftrightarrow \frac{\partial}{\partial x} F(x) = f(x)$$

La función de densidad verifica las dos propiedades siguientes:

i) $f(x) \geq 0$, $\forall x \in \mathbb{R}$ (salvo a lo sumo en una cantidad numerable de valores reales)

Demostración: Por reducción a lo absurdo, si existiera un conjunto A no numerable con $f(x) < 0$, $\forall x \in A$, entonces $P(A) = \int_A f(x)dx < 0$, lo cual es imposible.

ii) $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Demostración: Se verifica pues $\int_{-\infty}^{+\infty} f(x)dx = \int_{\mathbb{R}} f(x)dx = P(\Omega) = 1$

Ejemplo 2.6:

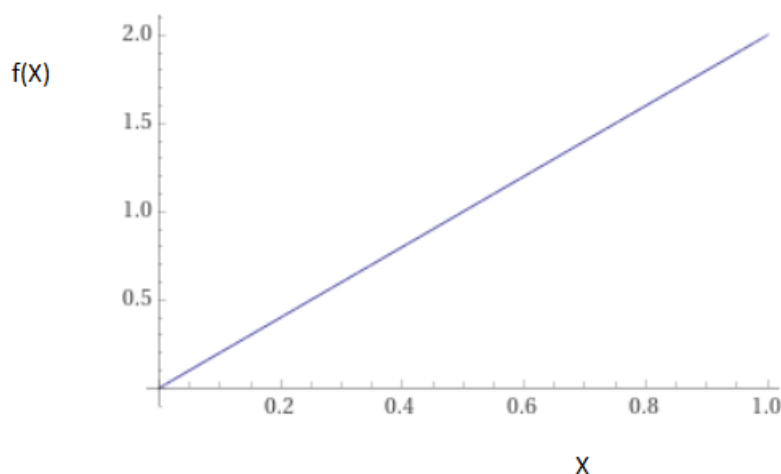
Sea X una variable aleatoria con densidad. $f(x) = \begin{cases} kx & \text{si } x \in [0, 1] \\ 0 & \text{fuera.} \end{cases}$

Determinar:

a) El valor de k

b) $P(X > 0,5)$

a) $\int_{\mathbb{R}} f(x) dx = 1 = \int_0^1 kx dx = \frac{k}{2}$, de donde $k = 2$.



b) $P(X > 0,5) = \int_{\frac{1}{2}}^1 2x dx = \frac{3}{4}$.

Distribuciones Mixtas

La mayoría de las distribuciones que se encuentran en problemas prácticos son discretas o continuas. Ahora se demostrará, sin embargo, que algunas veces puede ser necesario considerar una distribución que es una mezcla de una distribución discreta y una continua.

Supóngase que el voltaje de cierto sistema eléctrico es una variable aleatoria con una densidad dada por:

$$f(x) = \begin{cases} \frac{1}{(1+x)^2} & \text{para } x > 0 \\ 0 & \text{para } x \leq 0. \end{cases}$$

Pongámonos en la situación de que el voltaje de dicho sistema eléctrico se va a medir con un voltímetro que registra el valor real de X si $X \leq 3$, pero si $X > 3$, registra simplemente el valor 3. Si se define Y como el valor registrado por el voltímetro, entonces la distribución de Y se puede obtener como sigue:

En primer lugar, $P(Y = 3) = P(X \geq 3) = \frac{1}{4}$. Puesto que el valor $Y = 3$ tiene probabilidad $\frac{1}{4}$, resulta que $P(0 < Y < 3) = \frac{3}{4}$. Además, puesto que $Y = X$ para $0 < X < 3$, esta probabilidad $\frac{3}{4}$ para Y se distribuye sobre el intervalo $0 < Y < 3$ según la misma función de densidad con que la variable X se distribuye sobre el mismo intervalo. Entonces, la distribución de Y se especifica por la combinación de una función de densidad sobre el intervalo $0 < Y < 3$ y una probabilidad positiva en el punto $Y = 3$.

2.3 Modelos de probabilidad más relevantes

2.3.1 Distribucion de Bernoulli

En teoría de probabilidad y estadística, la distribución Bernoulli (o distribución dicotómica), es una distribución de probabilidad discreta, donde el valor 1 (éxito) ocurre con la probabilidad p y el valor 0 (fracaso) con la probabilidad $q = 1 - p$.

Un experimento al cual se aplica la distribución de Bernoulli se conoce como **ensayo de Bernoulli** o simplemente **ensayo** y la serie de esos experimentos independientes como **experimento de Bernoulli**.

Si se lleva a cabo un único experimento con dos posibles resultados denominados éxito y fracaso, se dice que la variable aleatoria X se distribuye como una Bernoulli de parámetro p con $0 < p < 1$ y escribimos $X \sim \text{Bernoulli}(p)$.

Función de probabilidad

Su función de probabilidad es

$$P(X = x) = p^x(1 - p)^{1-x} \quad x = 0, 1,$$

o equivalentemente,

$$P(X = x) = \begin{cases} 1 - p & \text{si } x = 0 \\ p & \text{si } x = 1 \end{cases},$$

que en ocasiones también suele escribirse como

$$P(X = x) = px + (1 - p)(1 - x) \quad x = 0, 1.$$

Función de distribución

La función de distribución acumulada de una variable aleatoria $X \sim \text{Bernoulli}(p)$ está dada por

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}.$$

Ejemplo 2.7: Se quiere lanzar una moneda equilibrada y se considera éxito obtener cruz. Si X es la variable dicotómica correspondiente, $X \sim \text{Bernoulli}(\frac{1}{2})$.

Ejemplo 2.8: Se lanza un dado equilibrado y se considera éxito obtener 6. Entonces, si X la variable dicotómica correspondiente, $X \sim \text{Bernoulli}(\frac{1}{6})$.

2.3.2 Distribucion de Binomial

En teoría de la probabilidad, la distribución binomial es una distribución de probabilidad discreta que cuenta el número de éxitos en una secuencia de n ensayos de Bernoulli independientes entre sí con una probabilidad fija p de ocurrencia de éxito en cada uno de los ensayos.

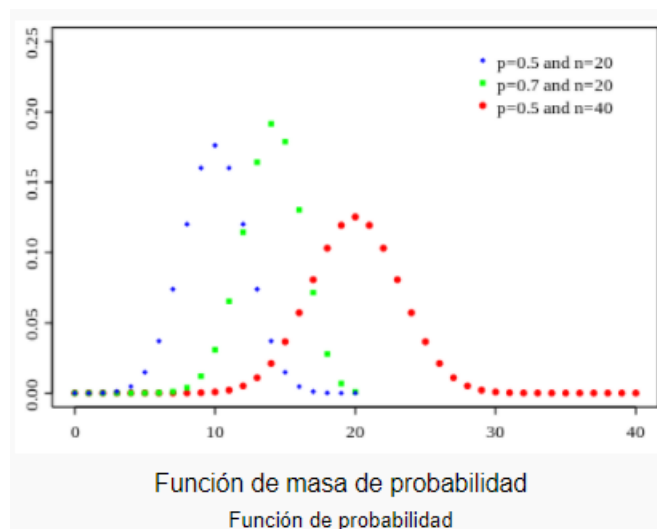
Sea X es una variable aleatoria discreta que mide el “número de éxitos” si se realizan n ensayos independientes de un experimento con dos posibles resultados denominados éxito y fracaso. Se dice que la variable aleatoria X se distribuye como una Binomial de parámetros n, p con $n > 0$ y $0 < p < 1$ y escribimos $X \sim \text{Bin}(n, p)$.

Función de probabilidad

Si $X \sim \text{Bin}(n, p)$ entonces su función de probabilidad está dada por

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

para $k = 0, 1, 2, \dots, n$, siendo $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ el coeficiente binomial.



Puede observarse que, como deber ser,

$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} = (p + 1 - p)^n = 1,$$

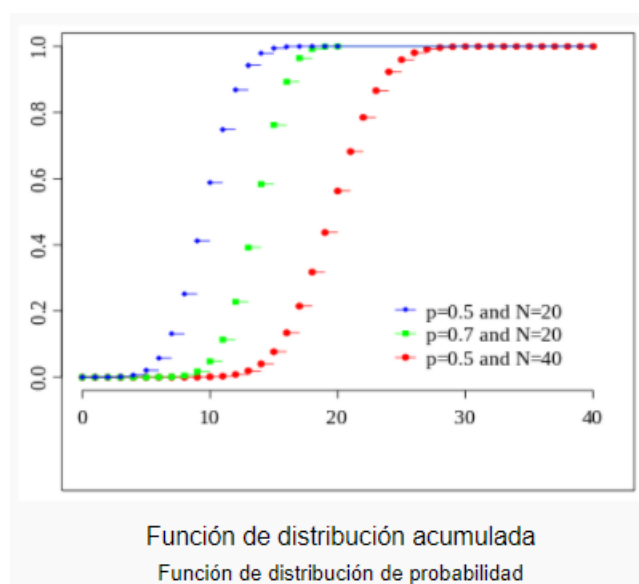
donde se ha hecho uso del binomio de Newton.

Función de distribución

La función de distribución acumulada de una variable aleatoria $X \sim \text{Bin}(n, p)$ está dada por

$$F(x) = P(X \leq x) = \begin{cases} \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k} & \text{si } x \geq 0 \\ 0 & \text{en otro caso,} \end{cases}$$

donde $\lfloor x \rfloor$ representa la parte entera de x , es decir, el mayor entero menor o igual que x .



Ejemplo 2.9:

Supongamos que se lanza 51 veces un dado de 6 caras equiprobables y queremos calcular la probabilidad de que el número 3 salga 20 veces.

En este problema un ensayo consiste en lanzar el dado una vez. Consideramos éxito si obtenemos un 3 pero si no sale 3 lo consideramos fracaso. Se define X como el número de veces que se obtiene un 3 en 51 lanzamientos.

En este caso tenemos $X \sim \text{Bin}(51, 1/6)$ por lo que la probabilidad buscada es $P(X = 20)$

$$P(X = 20) = \binom{51}{20} (1/6)^{20} (1 - 1/6)^{51-20} = 0,0000744.$$

2.3.3 Distribución geométrica

Volvamos a un experimento en el que se repiten de forma independiente ensayos de Bernoulli con probabilidad p de éxito. Sea X la variable aleatoria “número de ensayos hasta obtener el primer éxito”. Entonces, X es una variable aleatoria discreta con función de probabilidad dada por

$$P(X = n) = q^{n-1} p, \text{ para } n=1, 2, \dots$$

Observemos que se cumplen las condiciones de regularidad de toda función de probabilidad pues

$$\sum_{n=1}^{\infty} P(X = n) = \sum_{n=1}^{\infty} q^{n-1} p = p \sum_{n=1}^{\infty} q^{n-1} = p \frac{1}{1-q} = \frac{p}{p} = 1,$$

donde se ha hecho uso de la suma de una serie geométrica. A la distribución geométrica la notaremos $G(p)$. La función de distribución es

$$F(x) = \begin{cases} \sum_{n=1}^{\lfloor x \rfloor} q^{n-1} p & \text{si } x \geq 1 \\ 0 & \text{en otro caso,} \end{cases}$$

para $x \in \mathbb{R}$ donde $\lfloor x \rfloor$ representa la parte entera de x .

Ejemplo 2.10: La probabilidad de que una persona sea hincha del equipo de fútbol de una cierta ciudad es 0,6. Determinar la probabilidad de que si se pregunta aleatoriamente a los habitantes de dicha ciudad, el primero que responda que es hincha sea el cuarto encuestado.

X es una geométrica con $p = 0,6$, entonces

$$P(X = 4) = (0,4)^3 0,6.$$

2.3.4 Distribución binomial negativa

Situándonos, de nuevo, en los ensayos repetidos de Bernoulli con probabilidad p de éxito, la variable aleatoria discreta X que cuenta el número de fracasos hasta obtener el r -ésimo éxito sigue una distribución binomial negativa con función de probabilidad

$$P(X = k) = \binom{r+k-1}{r-1} q^k p^r, \text{ para } k=0, 1, 2, \dots$$

La denominación surge de que los números combinatorios $\binom{r+k-1}{r-1}$ admiten una notación (que aquí ignoraremos) en la que la parte superior el número correspondiente es negativo.

La probabilidad de k fracasos antes de r éxitos corresponde a cualquier sucesión de $k+r$ resultados que acabe en éxito y, por tanto, en los $r+k-1$ primeros ensayos deberán obtenerse $r-1$ éxitos. Hay $\binom{r+k-1}{r-1}$ posibilidades para tales resultados y cada uno (por independencia) tiene probabilidad $q^k p^{r-1} p$. A la distribución de X la notaremos $X \sim BN(r, p)$ y su función de distribución es

$$F(x) = \begin{cases} \sum_{k=0}^{\lfloor x \rfloor} \binom{r+k-1}{r-1} q^k p^r & \text{si } x \geq 0 \\ 0 & \text{en otro caso,} \end{cases}$$

2.3.5 Distribución Uniforme Discreta

La distribución uniforme discreta es una distribución de probabilidad que surge en espacios con puntos equiprobables, es decir, en situaciones donde los n resultados del experimento tienen la misma probabilidad de ocurrir.

Si X es una variable aleatoria discreta cuyo soporte es el conjunto $\{x_1, x_2, \dots, x_n\}$ y tiene una distribución uniforme discreta entonces escribiremos $X \sim U(x_1, x_2, \dots, x_n)$.

Función de Probabilidad

La función de probabilidad de X es

$$P(X = x) = \begin{cases} \frac{1}{n} & \text{para } x = x_1, x_2, \dots, x_n \\ 0 & \text{en otro caso.} \end{cases}$$

Función de distribución

Si $X \sim U(x_1, x_2, \dots, x_n)$ entonces la función de distribución es:

$$F(x) = \begin{cases} 0 & x < x_1 \\ \frac{i-1}{n} & x_{i-1} \leq x < x_i \\ 1 & x \geq x_n. \end{cases}$$

Ejemplo 2.11:

Para un dado perfecto, todos los resultados tienen la probabilidad de $\frac{1}{6}$ y, por tanto, la variable que mide el resultado del lanzamiento de dicho dado es $U(1, 2, 3, 4, 5, 6)$.

2.3.6 Distribución de Poisson

La distribución de Poisson es una distribución de probabilidad discreta que expresa, a partir de una frecuencia de ocurrencia media, la probabilidad de que aparezca un determinado número de eventos durante cierto período de tiempo. Concretamente, se especializa en la probabilidad de ocurrencia de sucesos con probabilidades muy pequeñas, o sucesos «raros».

La distribución de Poisson es popular porque modela el número de veces que ocurre un evento en un intervalo de tiempo.

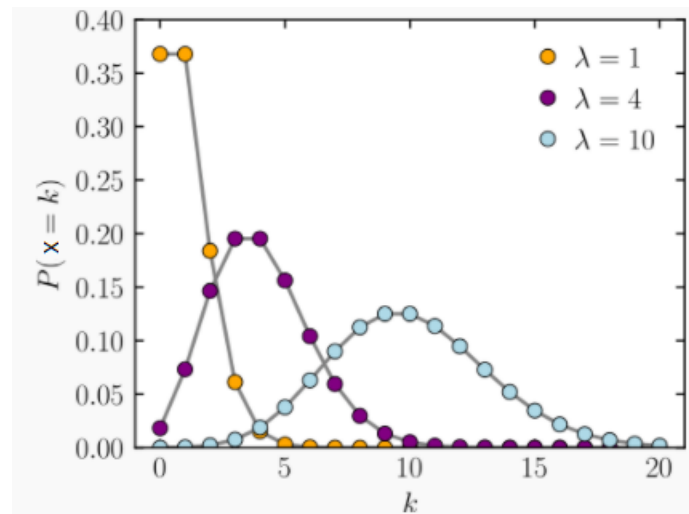
Sea $\lambda > 0$ y X una variable aleatoria discreta con distribución de Poisson de parámetro λ . Entonces, escribiremos $X \sim \text{Poisson}(\lambda)$ o $X \sim P(\lambda)$.

Función de Probabilidad

Si $X \sim \text{Poisson}(\lambda)$ entonces la función de probabilidad es

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \text{ donde } k = 0, 1, 2, \dots,$$

es el número de ocurrencias del evento o fenómeno.

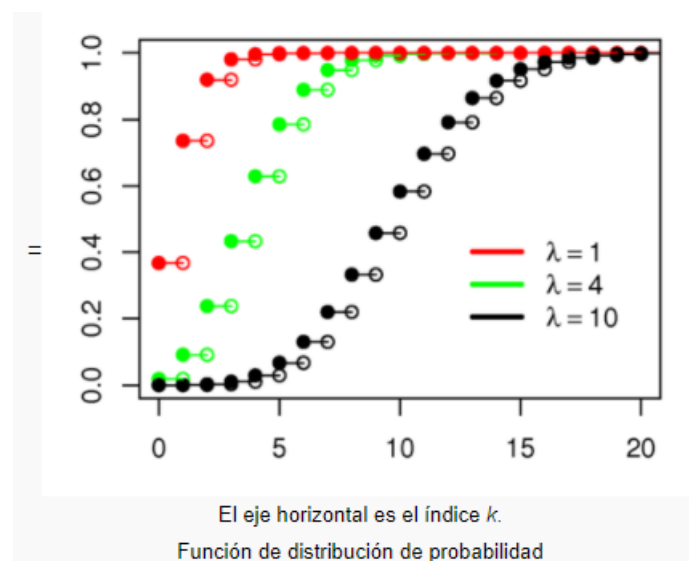


El parámetro $\lambda > 0$ representa el número de veces que se espera que ocurra el fenómeno durante un intervalo dado. Por ejemplo, si el suceso estudiado tiene lugar en promedio 4 veces por minuto y estamos interesados en la probabilidad de que ocurra k veces dentro de un intervalo de 10 minutos, usaremos un modelo de distribución de Poisson con $\lambda = 40$.

Función de distribución

La función de distribución vendrá dada por

$$F(x) = \begin{cases} \sum_{k=0}^{[x]} \frac{e^{-\lambda} \lambda^k}{k!} & \text{si } x \geq 0 \\ 0 & \text{en otro caso,} \end{cases}$$



Ejemplo 2.12: Si el 2% de los libros encuadernados en cierto taller tienen encuadernación defectuosa, para obtener la probabilidad de que 5 de 400 libros encuadernados en este taller tengan encuadernaciones defectuosas usamos la distribución de Poisson. Si se define X como el número de libros que tengan encuadernación defectuosa entonces $k = 5$ y λ (el valor esperado de libros defectuosos) es el 2% de 400, es decir, 8. Por lo tanto, la probabilidad buscada es:

$$P(X = 5) = \frac{8^5 e^{-8}}{5!} = 0,092$$

2.3.7 Distribución Uniforme Continua

Llamaremos así a una familia de distribuciones de probabilidad para variables aleatorias continuas, tales que para cada miembro de la familia, todos los intervalos de igual longitud en su rango (o dominio) son igualmente probables. El dominio está definido por dos parámetros, a y b , que son sus valores mínimo y máximo respectivamente.

La distribución o modelo uniforme puede considerarse como procedente de un proceso de extracción aleatoria. El planteamiento radica en el hecho de que la probabilidad se distribuye uniformemente a lo largo de un intervalo.

Función de densidad

Si X es una variable aleatoria con densidad uniforme en el intervalo (a, b) entonces la función de densidad es:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{para } x \in [a, b] \\ 0 & \text{en otro caso.} \end{cases}$$

Lo notaremos $X \sim U(a, b)$.

Función de distribución

Si $X \sim U(a, b)$ entonces la función de distribución es:

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b \end{cases}$$

la cual es fácil de obtener a partir de la función de densidad, pues para $a \leq x < b$

$$F(x) = \int_a^x \frac{1}{b-a} du = \frac{1}{b-a} u \Big|_a^x = \frac{x-a}{b-a}.$$

Ejemplo 2.13: Un autobús pasa por cierta parada cada 15 minutos. ¿Cuál es la probabilidad de que un señor que llega en un momento dado tenga que esperar al autobús más de 5 minutos?

Si se define X como el tiempo de espera, entonces $X \sim U(0, 15)$. Se calculará en primer lugar la función de distribución

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{15} & 0 \leq x < 15 \\ 1 & x \geq 15 \end{cases}$$

La probabilidad pedida viene dada por:

$$P(X > 5) = 1 - P(X \leq 5) = 1 - \frac{5}{15} = 0,67.$$

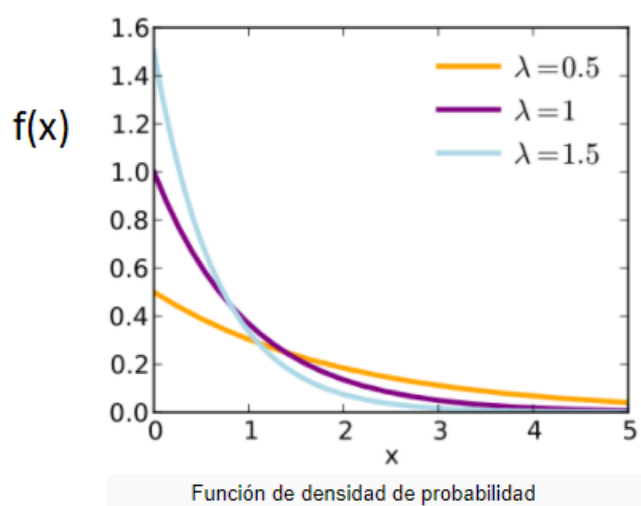
2.3.8 Distribución Exponencial

La distribución exponencial es una distribución continua que se utiliza para modelar tiempos de espera para la ocurrencia de un cierto evento. Esta distribución, al igual que la distribución geométrica, tiene la propiedad de pérdida de memoria. La distribución exponencial es un caso particular de la distribución gamma, que veremos más adelante.

Función de densidad

Se dice que una variable aleatoria continua X tiene una distribución exponencial con parámetro $\lambda > 0$ y escribimos $X \sim \text{Exp}(\lambda)$ si su función de densidad es

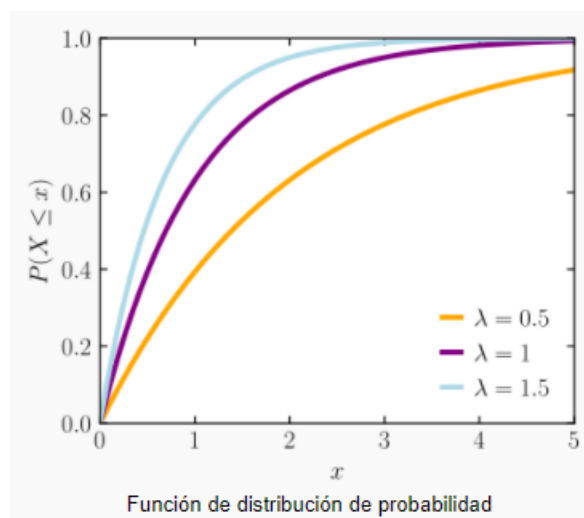
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$



Función de distribución

Su función de distribución acumulada está dada por

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$



La propiedad de falta de memoria de la distribución exponencial afirma que, si $t, s > 0$,

$$P(X > s+t | X > s) = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t).$$

Es decir, si la duración de un aparato electrónico sigue una distribución exponencial de parámetro λ la probabilidad de que dure t instantes de tiempo más cuando lleva funcionando s es la misma que la probabilidad de que dure t instantes de tiempo más desde el principio.

Ejemplo 2.14: Se ha comprobado que el tiempo de vida de cierto tipo de marcapasos sigue una distribución exponencial con $\lambda = \frac{1}{16}$. ¿Cuál es la probabilidad de que a una persona a la que se le ha implantado este marcapasos se le deba reimplantar otro antes de 20 años? Si el marcapasos lleva funcionando correctamente 5 años en un paciente, ¿Cuál es la probabilidad de que haya que cambiarlo antes de 25 años?

Sea X la variable aleatoria que mide la duración de un marcapasos en una persona. Tenemos que $X \sim \text{Exp}(\lambda = \frac{1}{16})$, con función de distribución, para $x > 0$,

$$F(x) = 1 - e^{-\frac{1}{16}x}.$$

Entonces

$$P(X \leq 20) = F(20) = 1 - e^{-\frac{20}{16}} = 0,7135.$$

En segundo lugar,

$$P(X \leq 25 | X > 5) = \frac{P(5 < X \leq 25)}{P(X > 5)} = \frac{F(25) - F(5)}{1 - F(5)} = \frac{1 - e^{-\frac{25}{16}} - 1 + e^{-\frac{5}{16}}}{e^{-\frac{5}{16}}} = 0,7135.$$

2.3.9 Distribución Gamma

La distribución gamma es una distribución con dos parámetros que pertenece a las distribuciones de probabilidad continuas. La distribución exponencial, distribución de Erlang y la distribución χ^2 son casos particulares de la distribución gamma.

Si una variable aleatoria continua X tiene distribución gamma con parámetros $\alpha > 0$ y $\lambda > 0$ entonces escribiremos $X \sim \Gamma(\alpha, \lambda)$.

Función de densidad

Si $X \sim \Gamma(\alpha, \lambda)$ entonces su función de densidad es

$$f(x) = \begin{cases} \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0, \end{cases}$$

donde

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt,$$

es la función gamma y satisface

- $\Gamma(2) = \Gamma(1) = 1$
- Para cualquier $\alpha > 0$ se cumple que $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
- Si $n \in \mathbb{Z}^+$ entonces $\Gamma(n + 1) = n!$

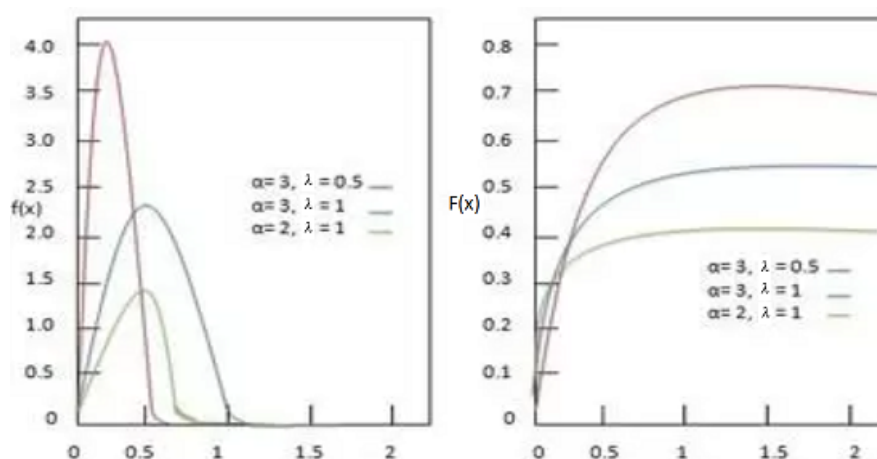
Función de distribución

La función de distribución acumulada de una variable aleatoria $X \sim \Gamma(\alpha, \lambda)$ está dada por

$$F(x) = \begin{cases} \int_0^x \frac{\lambda(\lambda y)^{\alpha-1} e^{-\lambda y}}{\Gamma(\alpha)} dy & \text{si } x > 0 \\ 0 & \text{si } x \leq 0. \end{cases}$$

Si X es una variable aleatoria tal que $X \sim \Gamma(n, \lambda)$ donde $n \in \mathbb{Z}^+$ (es decir, X tiene una distribución de Erlang) entonces su función de distribución acumulada está dada por

$$F(x) = \begin{cases} 1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} e^{-\lambda x} = \sum_{k=n}^{\infty} \frac{(\lambda x)^k}{k!} e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0. \end{cases}$$



La distribución $\Gamma(n, \lambda)$ mide el tiempo de espera hasta la n -ésima llegada en un proceso de Poisson con λ llegadas en la unidad de tiempo.

2.3.10 Distribución Normal

En estadística y probabilidad se llama distribución normal, distribución de Gauss, distribución gaussiana o distribución de Laplace-Gauss, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en estadística y en la teoría de probabilidades.

La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro estadístico. Esta curva se conoce como campana de Gauss y es el gráfico de una función gaussiana.

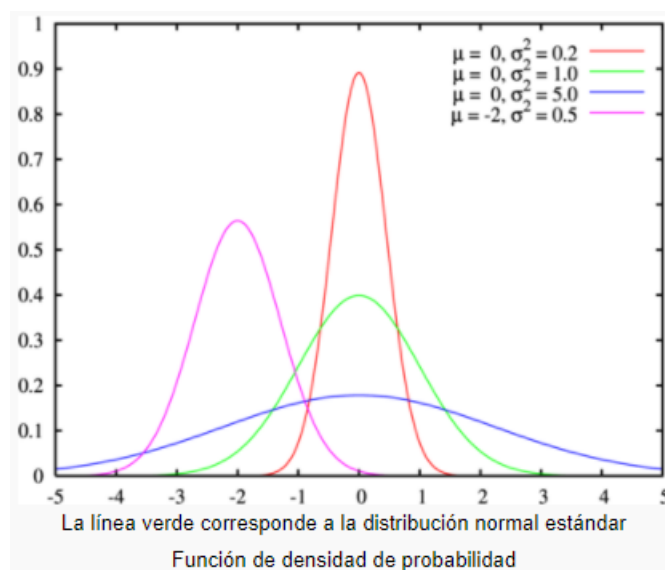
La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos. Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la enorme cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.

Función de densidad

Si $X \sim N(\mu, \sigma)$ entonces su función de densidad es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ si } x \in \mathbb{R}.$$

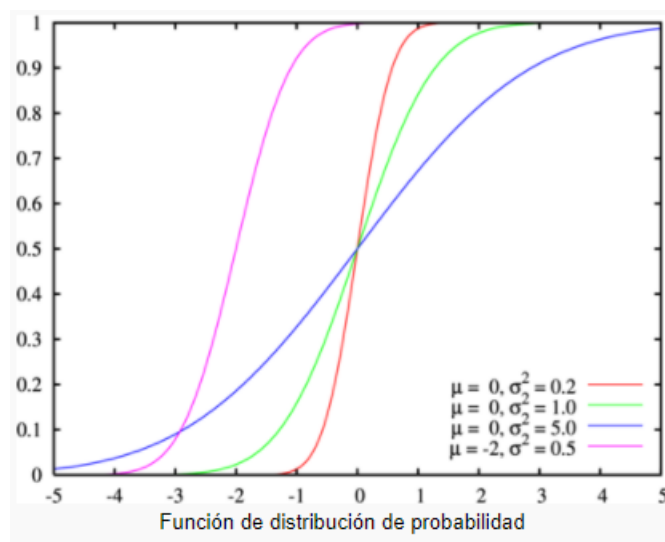
Cuando $\mu = 0$ y $\sigma = 1$, diremos que la normal es tipificada y lo simbolizaremos por $N(0, 1)$ o muy comúnmente por Z . Si $X \sim N(\mu, \sigma)$, entonces $\frac{X-\mu}{\sigma} \sim N(0, 1) \equiv Z$.



Función de distribución

La función de distribución de la normal está definida como sigue:

$$F(x) = \int_{-\infty}^x f(u) du = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(u-\mu)^2}{2\sigma^2}} du \text{ si } x \in \mathbb{R}.$$

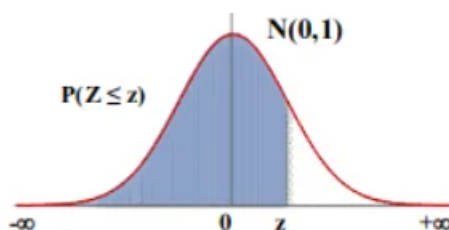


No existe primitiva para la función de densidad normal y los cálculos se hacen numéricamente (o con ordenador) estando tabulados los resultados.

Ejemplos : Dada una variable aleatoria $X \sim N(4, 2)$ se pide calcular $P(X > 3)$.

$$P(X > 3) = P\left(\frac{X-4}{2} > \frac{3-4}{2}\right) = P\left(Z > -\frac{1}{2}\right) = P(Z < 0,5) = 0,6915,$$

valor que se obtiene de la tabla de la $N(0, 1)$.



| z | 0,00 | 0,01 | 0,02 | 0,03 | 0,04 | 0,05 | 0,06 | 0,07 | 0,08 | 0,09 |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 0,0 | 0,5000 | 0,5040 | 0,5080 | 0,5120 | 0,5160 | 0,5199 | 0,5239 | 0,5279 | 0,5319 | 0,5359 |
| 0,1 | 0,5398 | 0,5438 | 0,5478 | 0,5517 | 0,5557 | 0,5596 | 0,5636 | 0,5675 | 0,5714 | 0,5753 |
| 0,2 | 0,5793 | 0,5832 | 0,5871 | 0,5910 | 0,5948 | 0,5987 | 0,6026 | 0,6064 | 0,6103 | 0,6141 |
| 0,3 | 0,6179 | 0,6217 | 0,6255 | 0,6293 | 0,6331 | 0,6368 | 0,6406 | 0,6443 | 0,6480 | 0,6517 |
| 0,4 | 0,6554 | 0,6591 | 0,6628 | 0,6664 | 0,6700 | 0,6736 | 0,6772 | 0,6808 | 0,6844 | 0,6879 |
| 0,5 | 0,6915 | 0,6950 | 0,6985 | 0,7019 | 0,7054 | 0,7088 | 0,7123 | 0,7157 | 0,7190 | 0,7224 |
| 0,6 | 0,7257 | 0,7291 | 0,7324 | 0,7357 | 0,7389 | 0,7422 | 0,7454 | 0,7486 | 0,7517 | 0,7549 |
| 0,7 | 0,7580 | 0,7611 | 0,7642 | 0,7673 | 0,7704 | 0,7734 | 0,7764 | 0,7794 | 0,7823 | 0,7852 |
| 0,8 | 0,7881 | 0,7910 | 0,7939 | 0,7967 | 0,7995 | 0,8023 | 0,8051 | 0,8078 | 0,8106 | 0,8133 |
| 0,9 | 0,8159 | 0,8186 | 0,8212 | 0,8238 | 0,8264 | 0,8289 | 0,8315 | 0,8340 | 0,8365 | 0,8389 |
| 1,0 | 0,8413 | 0,8438 | 0,8461 | 0,8485 | 0,8508 | 0,8531 | 0,8554 | 0,8577 | 0,8599 | 0,8621 |
| 1,1 | 0,8643 | 0,8665 | 0,8686 | 0,8708 | 0,8729 | 0,8749 | 0,8770 | 0,8790 | 0,8810 | 0,8830 |
| 1,2 | 0,8849 | 0,8869 | 0,8888 | 0,8907 | 0,8925 | 0,8944 | 0,8962 | 0,8980 | 0,8997 | 0,9015 |
| 1,3 | 0,9032 | 0,9049 | 0,9066 | 0,9082 | 0,9099 | 0,9115 | 0,9131 | 0,9147 | 0,9162 | 0,9177 |
| 1,4 | 0,9192 | 0,9207 | 0,9222 | 0,9236 | 0,9251 | 0,9265 | 0,9279 | 0,9292 | 0,9306 | 0,9319 |
| 1,5 | 0,9332 | 0,9345 | 0,9357 | 0,9370 | 0,9382 | 0,9394 | 0,9406 | 0,9418 | 0,9429 | 0,9441 |
| 1,6 | 0,9452 | 0,9463 | 0,9474 | 0,9484 | 0,9495 | 0,9505 | 0,9515 | 0,9525 | 0,9535 | 0,9545 |
| 1,7 | 0,9554 | 0,9564 | 0,9573 | 0,9582 | 0,9591 | 0,9599 | 0,9608 | 0,9616 | 0,9625 | 0,9633 |
| 1,8 | 0,9641 | 0,9649 | 0,9656 | 0,9664 | 0,9671 | 0,9678 | 0,9686 | 0,9693 | 0,9699 | 0,9706 |
| 1,9 | 0,9713 | 0,9719 | 0,9726 | 0,9732 | 0,9738 | 0,9744 | 0,9750 | 0,9756 | 0,9761 | 0,9767 |
| 2,0 | 0,9772 | 0,9778 | 0,9783 | 0,9788 | 0,9793 | 0,9798 | 0,9803 | 0,9808 | 0,9812 | 0,9817 |
| 2,1 | 0,9821 | 0,9826 | 0,9830 | 0,9834 | 0,9838 | 0,9842 | 0,9846 | 0,9850 | 0,9854 | 0,9857 |
| 2,2 | 0,9861 | 0,9864 | 0,9868 | 0,9871 | 0,9875 | 0,9878 | 0,9881 | 0,9884 | 0,9887 | 0,9890 |
| 2,3 | 0,9893 | 0,9896 | 0,9898 | 0,9901 | 0,9904 | 0,9906 | 0,9909 | 0,9911 | 0,9913 | 0,9916 |
| 2,4 | 0,9918 | 0,9920 | 0,9922 | 0,9925 | 0,9927 | 0,9929 | 0,9931 | 0,9932 | 0,9934 | 0,9936 |
| 2,5 | 0,9938 | 0,9940 | 0,9941 | 0,9943 | 0,9945 | 0,9946 | 0,9948 | 0,9949 | 0,9951 | 0,9952 |
| 2,6 | 0,9953 | 0,9955 | 0,9956 | 0,9957 | 0,9959 | 0,9960 | 0,9961 | 0,9962 | 0,9963 | 0,9964 |
| 2,7 | 0,99653 | 0,99664 | 0,99674 | 0,99683 | 0,99693 | 0,99702 | 0,99711 | 0,99720 | 0,99728 | 0,99736 |
| 2,8 | 0,99744 | 0,99752 | 0,99760 | 0,99767 | 0,99774 | 0,99781 | 0,99788 | 0,99795 | 0,99801 | 0,99807 |
| 2,9 | 0,99813 | 0,99819 | 0,99825 | 0,99831 | 0,99836 | 0,99841 | 0,99846 | 0,99851 | 0,99856 | 0,99861 |
| 3,0 | 0,99865 | 0,99869 | 0,99874 | 0,99878 | 0,99882 | 0,99886 | 0,99889 | 0,99893 | 0,99896 | 0,99900 |
| 3,1 | 0,99903 | 0,99906 | 0,99910 | 0,99913 | 0,99916 | 0,99918 | 0,99921 | 0,99924 | 0,99926 | 0,99929 |
| 3,2 | 0,99931 | 0,99934 | 0,99936 | 0,99938 | 0,99940 | 0,99942 | 0,99944 | 0,99946 | 0,99948 | 0,99950 |
| 3,3 | 0,99952 | 0,99953 | 0,99955 | 0,99957 | 0,99958 | 0,99960 | 0,99961 | 0,99962 | 0,99964 | 0,99965 |
| 3,4 | 0,99966 | 0,99968 | 0,99969 | 0,99970 | 0,99971 | 0,99972 | 0,99973 | 0,99974 | 0,99975 | 0,99976 |
| 3,5 | 0,99977 | 0,99978 | 0,99978 | 0,99979 | 0,99980 | 0,99981 | 0,99981 | 0,99982 | 0,99983 | 0,99983 |
| 3,6 | 0,99984 | 0,99985 | 0,99985 | 0,99986 | 0,99986 | 0,99987 | 0,99987 | 0,99988 | 0,99988 | 0,99989 |
| 3,7 | 0,99989 | 0,99990 | 0,99990 | 0,99990 | 0,99991 | 0,99991 | 0,99992 | 0,99992 | 0,99992 | 0,99992 |
| 3,8 | 0,99993 | 0,99993 | 0,99993 | 0,99994 | 0,99994 | 0,99994 | 0,99994 | 0,99995 | 0,99995 | 0,99995 |
| 3,9 | 0,99995 | 0,99995 | 0,99996 | 0,99996 | 0,99996 | 0,99996 | 0,99996 | 0,99996 | 0,99997 | 0,99997 |
| 4,0 | 0,99997 | 0,99997 | 0,99997 | 0,99997 | 0,99997 | 0,99997 | 0,99998 | 0,99998 | 0,99998 | 0,99998 |

Ejemplos y/o aplicaciones:

Algunos ejemplos de variables asociadas a fenómenos naturales que siguen el modelo de la normal son:

- Caracteres morfológicos de individuos como la estatura.
- Caracteres fisiológicos como el efecto de un fármaco.
- Caracteres sociológicos como el consumo de cierto producto por un mismo grupo de

individuos.

- Caracteres psicológicos como el cociente intelectual.
- Nivel de ruido en telecomunicaciones.
- Errores cometidos al medir ciertas magnitudes.

2.3.11 Distribuciones Beta

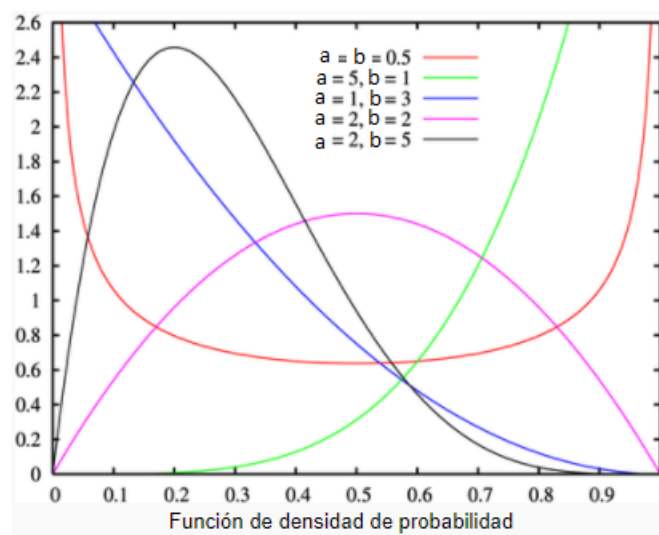
Las distribuciones beta constituyen una familia de distribuciones de probabilidad continuas definidas en el intervalo $(0, 1)$ parametrizadas por dos parámetros positivos, denotados por a y b , que controlan la forma de la distribución.

Si una variable aleatoria continua X tiene una distribución beta con parámetros $a, b > 0$ entonces escribiremos $X \sim B(a, b)$.

Función de densidad

La función de densidad de X es

$$f(x) = \begin{cases} \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} & \text{si } x \in [0, 1] \\ 0 & \text{si } x \notin [0, 1]. \end{cases}$$



Donde $B(a, b)$ es la función beta y se define para $a, b > 0$ como

$$B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx$$

y algunas de las propiedades que satisface son:

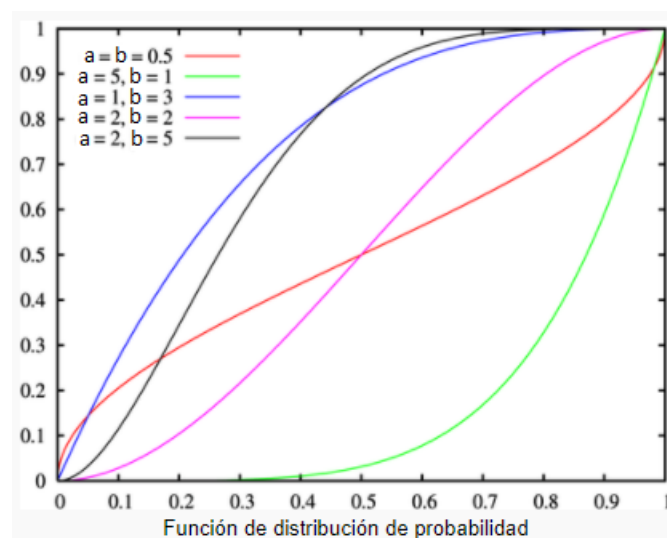
- $B(a, b) = B(b, a)$

$$\bullet B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Función de distribución

La función de distribución de X es

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{\int_0^x t^{a-1} (1-t)^{b-1} dt}{B(a, b)} & \text{si } x \in [0, 1] \\ 1 & \text{si } x > 1. \end{cases}$$



El caso particular $a = 1, b = 1$ es la distribución $U(0, 1)$. El máximo de n variables independientes todas con distribución $U(0, 1)$ es una $B(n, 1)$. El mínimo de n variables independientes todas con distribución $U(0, 1)$ es una $B(1, n)$. La familia Beta tiene muchas aplicaciones en la inferencia bayesiana del parámetro p de una distribución Bernoulli.

Bibliografía

1. A.A. Borovkov (2013). *Probability Theory*. Springer.
2. A. Klenke (2014). *Probability Theory. A Comprehensive Course*. Springer.
3. M.H. DeGroot (1988). *Probability and Statistics*. Wilmington, Delaware, E.U.A. Addison-Wesley Iberoamericana, S.A.
4. V.K. Rohatgi and A.K. Md. Ehsanes Saleh (2001). *An Introduction to Probability and Statistics*. John Wiley and Sons.

Tema 3

Esperanza matemática. Propiedades. Varianza. Propiedades. Función característica y función generatriz de momentos. Acotación de Tchebychev

3.1 Esperanza matemática.

3.1.1 Esperanza de una distribución discreta

Definición 3.1:

Dada una variable aleatoria X discreta cuya función de probabilidad es $P(X = i)$, $i \in I$, entonces, la esperanza matemática de X , que se denota $E(X)$, se define como sigue:

$$E(X) = \sum_{i \in I} iP(X = i) \quad (3.1)$$

Ejemplo 3.1:

Supóngase que una variable aleatoria X puede tomar únicamente los valores $-2, 0, 1$ y 4 y que $P(X = -2) = 0,1$, $P(X = 0) = 0,4$, $P(X = 1) = 0,3$ y $P(X = 4) = 0,2$. Entonces

$$E(X) = -2(0,1) + 0(0,4) + 1(0,3) + 4(0,2) = 0,9.$$

Se puede observar que la esperanza $E(X)$ no es necesariamente igual a ninguno de los valores posibles de X .

Si X puede tomar únicamente un número finito de valores distintos, como en el ejemplo anterior, entonces existe únicamente un número finito de términos en la suma de la ecuación 3.1. Sin embargo, si existe una sucesión infinita de valores posibles distintos de X , entonces la suma de la ecuación 3.1 es una serie infinita de términos. tal serie puede no converger para una función de probabilidad concreta. Se dice que la esperanza $E(X)$ existe si, y sólo si, la suma de la ecuación 3.1 es absolutamente convergente, esto es, si, y sólo si,

$$\sum_i |i|P(X = i) < \infty. \quad (3.2)$$

En otras palabras, si se verifica la relación 3.2, entonces $E(X)$ existe y su valor está dado por la ecuación 3.1. Si no se verifica la relación 3.2, entonces $E(X)$ no existe.

Ejemplo 3.2:

Se lanzan dos dados. Sea X el mínimo de los puntos obtenidos en ambos dados.

| X | P(X=i) |
|---|--------|
| 1 | 11/36 |
| 2 | 9/36 |
| 3 | 7/36 |
| 4 | 5/36 |
| 5 | 3/36 |
| 6 | 1/36 |

$$E(X) = 1 \cdot \frac{11}{36} + 2 \cdot \frac{9}{36} + \dots + 6 \cdot \frac{1}{36} = \frac{9}{4}$$

Ejemplo 3.3:

Se lanzan dos dados. Sea X el máximo de los puntos obtenidos en ambos dados.

| X | P(X=i) |
|---|--------|
| 1 | 1/36 |
| 2 | 3/36 |
| 3 | 5/36 |
| 4 | 7/36 |
| 5 | 9/36 |
| 6 | 11/36 |

$$E(X) = 1 \cdot \frac{1}{36} + 2 \cdot \frac{3}{36} + \dots + 6 \cdot \frac{11}{36} = \frac{161}{36}$$

Ejemplo 3.4:

Un jugador lanza un dado corriente. Si sale 1 o número primo, gana tantos cientos de euros como marca el dado, pero si no sale número primo, pierde tantos cientos de euros como marca el dado. Determinar la función de probabilidad y la esperanza matemática del juego.

| X | P(X=i) |
|------|--------|
| 100 | 1/6 |
| 200 | 1/6 |
| 300 | 1/6 |
| -400 | 1/6 |
| 500 | 1/6 |
| -600 | 1/6 |

$$E(X) = 100 \cdot \frac{1}{6} + 200 \cdot \frac{1}{6} + \dots + (-600) \cdot \frac{1}{6} = \frac{50}{3} = 16,66667$$

Ejemplo 3.5:

Calcular la media del número de bolas blancas obtenidas cuando se extraen simultáneamente dos bolas de una urna que contiene 5 blancas y tres negras, todas distinguibles. Si denotamos X a la variable que define el número de bolas blancas obtenidas al extraer las dos bolas, su función de probabilidad es: $P(X=0) = \frac{6}{56}$, $P(X=1) = \frac{30}{56}$ y $P(X=2) = \frac{20}{56}$. Ya que esta variable sólo toma tres valores, existe su esperanza y vale:

$$E(X) = 0 \cdot \frac{6}{56} + 1 \cdot \frac{30}{56} + 2 \cdot \frac{20}{56} = \frac{5}{4} = 1,25.$$

Así, el número medio de bolas blancas extraídas es 1.25, valor comprendido en el rango de valores de X (el intervalo $[0, 2]$), que resulta al ponderar cada valor de X por su probabilidad.

3.1.2 Esperanza de una distribución continua

Definición 3.2:

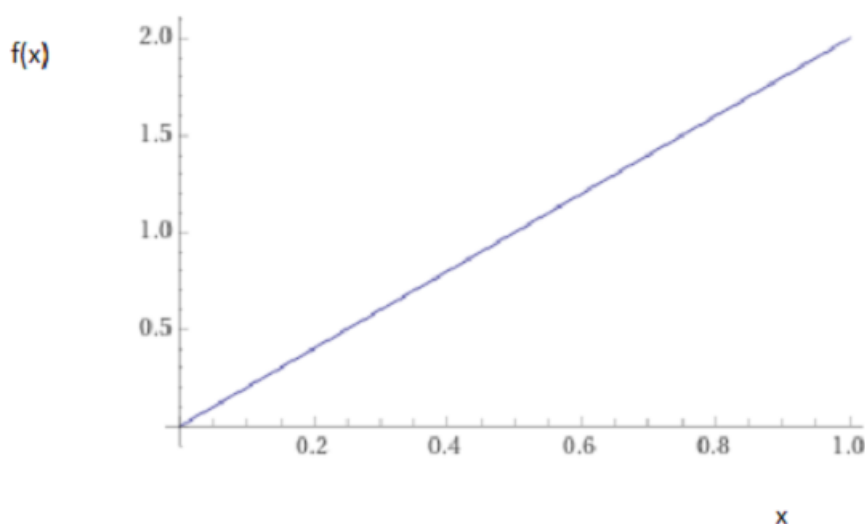
Sea una variable aleatoria X con distribución absolutamente continua cuya función de densidad es f . Entonces, la esperanza matemática de X , que se denota $E(X)$, se define como sigue:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx. \quad (3.3)$$

Ejemplo 3.6:

Supóngase que la función de densidad de una variable aleatoria X es

$$f(x) = \begin{cases} 2x & \text{si } x \in (0, 1) \\ 0 & \text{fuera.} \end{cases}$$



Entonces,

$$E(X) = \int_0^1 x \cdot f(x) dx = \int_0^1 x(2x) dx = \frac{2}{3}$$

Se dice que la esperanza $E(X)$ existe para una distribución continua si, y sólo si, la integral de la ecuación 3.3 es absolutamente convergente, esto es, si, y sólo si,

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty.$$

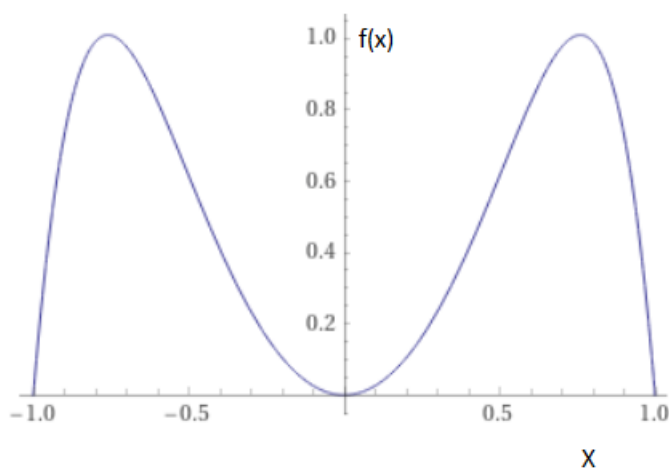
Siempre que X sea una variable aleatoria acotada, esto es, siempre que existan números a y b ($-\infty < a < b < \infty$) tales que $P(a \leq X \leq b) = 1$, como en el ejemplo anterior, entonces $E(X)$

debe existir.

Ejemplo 3.7:

Sea X una variable aleatoria con densidad:

$$f(x) = \begin{cases} \frac{21}{8}x^2(1-x^4) & \text{si } x \in [-1, 1] \\ 0 & \text{fuera.} \end{cases}$$

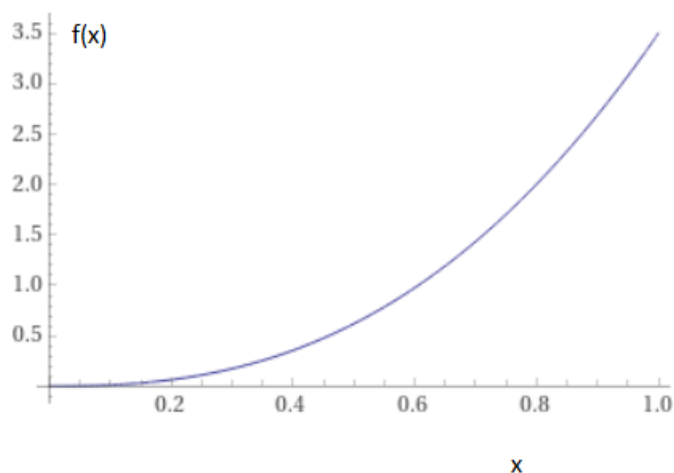


$$E(X) = \int x \cdot f(x) dx = \int_{-1}^1 x \left(\frac{21}{8} x^2 (1 - x^4) \right) dx = 0$$

Ejemplo 3.8:

Sea X una variable aleatoria con densidad.

$$f(x) = \begin{cases} \frac{7}{2}x^2\sqrt{x} & \text{si } x \in [0, 1] \\ 0 & \text{fuera.} \end{cases}$$



$$E(X) = \int x \cdot f(x) dx = \int_0^1 x \left(\frac{7}{2} x^2 \sqrt{x} \right) dx = \frac{7}{9}$$

3.2 Esperanza de una función

Si X es una variable aleatoria cuya función de densidad es f , entonces la esperanza de cualquier función $r(X)$ se puede determinar aplicando la definición de esperanza a la distribución de $r(X)$ como sigue: Se considera $Y = r(X)$; se determina la distribución de probabilidad de Y ; y entonces se determina $E(Y)$ aplicando la ecuación 3,1 o la 3,3.

Sin embargo, no es realmente necesario determinar la función de densidad de $r(X)$ para calcular la esperanza $E[r(X)]$. De hecho, se puede demostrar que el valor de $E[r(X)]$ siempre se puede calcular directamente a partir de la relación

$$E[r(X)] = \int_{-\infty}^{\infty} r(x) f(x) dx. \quad (3.4)$$

La esperanza de $E[r(X)]$ existe si, y sólo si,

$$\int_{-\infty}^{\infty} |r(x)| f(x) dx < \infty.$$

Si la distribución de Y fuera discreta, la integral de la ecuación 3,4 sería reemplazada por una suma.

3.3 Propiedades de la Esperanza

Supóngase que X es una variable aleatoria cuya esperanza $E(X)$ existe. Se presentarán varios resultados relacionados con las propiedades básicas de la esperanza.

i) Si $Y = aX + b$, donde a y b son constantes, entonces

$$E(Y) = aE(X) + b$$

Demostración:

En primer lugar, supóngase, por conveniencia, que X tiene una distribución continua cuya función de densidad es f . Entonces,

$$E(Y) = E(aX + b) = \int_{-\infty}^{\infty} (ax + b)f(x)dx = a \int_{-\infty}^{\infty} xf(x)dx + b \int_{-\infty}^{\infty} f(x)dx = aE(X) + b.$$

Se puede hacer una demostración análoga para una distribución discreta o para un tipo de distribución más general.

Ejemplo 3.9:

Sea X una variable aleatoria con densidad.

$$f(x) = \begin{cases} \frac{7}{2}x^2\sqrt{x} & \text{si } x \in [0, 1] \\ 0 & \text{fuera.} \end{cases}$$

Como se vio,

$$E(X) = \int x \cdot f(x)dx = \int_0^1 x \left(\frac{7}{2}x^2\sqrt{x} \right) dx = \frac{7}{9}.$$

Entonces

$$E(3X) = 3E(X) = 3\left(\frac{7}{9}\right) = \frac{7}{3}$$

Ejemplo 3.10:

Sea X una variable aleatoria con densidad:

$$f(x) = \begin{cases} \frac{21}{8}x^2(1-x^4) & \text{si } x \in [-1, 1] \\ 0 & \text{fuera.} \end{cases}$$

$$E(X) = \int x \cdot f(x)dx = \int_{-1}^1 x \left(\frac{21}{8}x^2(1-x^4) \right) dx = 0.$$

Entonces

$$E(2X + 3) = 2E(X) + 3 = 3$$

ii) Si existe una constante a tal que $P(X \geq a) = 1$, entonces $E(X) \geq a$. Si existe una constante b tal que $P(X \leq b) = 1$, entonces $E(X) \leq b$.

Demostración:

Supóngase de nuevo, por conveniencia, que X tiene una distribución continua cuya función de densidad es f y supóngase en primer lugar que $P(X \geq a) = 1$. Entonces,

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_a^{\infty} xf(x)dx \geq \int_a^{\infty} af(x)dx = aP(X \geq a) = a.$$

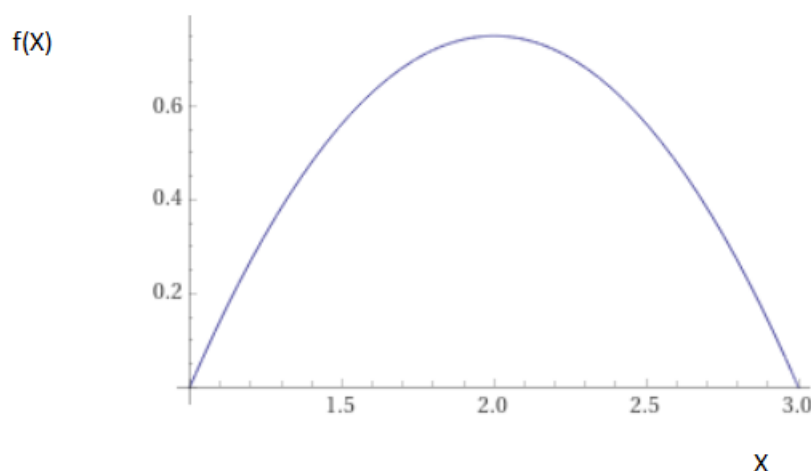
La demostración de la otra propiedad y la demostración para una distribución discreta o un tipo de distribución más general son análogas.

Por la propiedad ii) resulta que si $P(a \leq X \leq b) = 1$, entonces $a \leq E(X) \leq b$. Además se puede demostrar que si $P(X \geq a) = 1$ y si $E(X) = a$, entonces se debe verificar que $P(X > a) = 0$ y $P(X = a) = 1$.

Ejemplo 3.11:

La longitud de ciertos tornillos (en centímetros) es una variable aleatoria con la siguiente función de densidad:

$$f(x) = \begin{cases} \frac{3}{4}(-x^2 + 4x - 3) & \text{si } 1 \leq x \leq 3 \\ 0 & \text{fuera.} \end{cases}$$



Si para construir lo que se necesita con uno de estos tornillos hay que hacer un gasto de 10 euros por centímetro de longitud que tenga el tornillo más un gasto fijo de 4 euros, ¿cuál es el gasto medio esperado por tornillo?

La variable gasto G depende de la variable X de la siguiente forma: $G = 10X + 4$. Aplicando el operador esperanza a cada miembro y usando sus propiedades obtenemos:

$$E(G) = 10E(X) + 4$$

Entonces debemos calcular $E(X)$,

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_1^3 x \frac{3}{4}(-x^2 + 4x - 3) = 2$$

Notemos que la función de densidad es simétrica respecto de $x = 2$. Así que es razonable que hallamos obtenido que $E(X) = 2$. Entonces:

$$E(G) = 10(2) + 4 = 24.$$

3.4 Varianza

Supóngase que X es una variable aleatoria con media $\mu = E(X)$. La varianza de X , que se denotará por $Var(X)$, se define como sigue:

$$Var(X) = E[(X - \mu)^2] \quad (3.5)$$

Como $Var(X)$ es el valor esperado de la variable aleatoria no negativa $(X - \mu)^2$, resulta que $Var(X) \geq 0$. Se debe recordar que la esperanza de la ecuación 3,5 puede ser, o no, infinita. Si la esperanza no es finita, se dice que la $Var(X)$ no existe. Sin embargo, si los valores posibles de X están acotados, entonces $Var(X)$ debe existir.

El valor de $Var(X)$ se llama también varianza de la distribución X . Por tanto, en algunos casos se habla de la varianza de X y en otros de la varianza de una distribución de probabilidad.

La varianza de una distribución proporciona una medida de la variación o dispersión de una distribución alrededor de su media μ . Un valor pequeño de la varianza indica que la distribución de probabilidad está muy concentrada alrededor de μ ; y un valor grande de la varianza generalmente indica que la distribución de probabilidad tiene una dispersión amplia alrededor de μ . Sin embargo, la varianza de cualquier distribución, así como su media, se pueden hacer arbitrariamente grandes colocando una masa de probabilidad positiva, aunque sea pequeña, suficientemente lejos del origen en la recta real.

La desviación típica de una variable aleatoria o de una distribución se define como la raíz cuadrada no negativa de la varianza. La desviación típica, también llamada desviación estándar, de una variable aleatoria se denota usualmente por el símbolo σ y la varianza se denota por σ^2 .

Ejemplo 3.12:

En un casino existe un dado de 50 caras, la siguiente tabla recoge los posibles números que aparecen en cada una de las caras y la cantidad de caras con dicho número.

| Número | Cantidad |
|--------|----------|
| 9 | 1 |
| 10 | 4 |
| 11 | 9 |
| 12 | 16 |
| 13 | 11 |
| 14 | 8 |
| 15 | 1 |

Se lanza el dado una vez y sea X el número que sale en la cara superior. Se pide calcular la varianza.

Calculamos ahora la Esperanza matemática $E(X)$ y $Var(X)$ como:

$$E(X) = \frac{610}{50} = 12,2$$

$$\begin{aligned} Var(X) &= (9 - 12,2)^2 \frac{1}{50} + (10 - 12,2)^2 \frac{4}{50} + (11 - 12,2)^2 \frac{9}{50} + (12 - 12,2)^2 \frac{16}{50} \\ &\quad + (13 - 12,2)^2 \frac{11}{50} + (14 - 12,2)^2 \frac{8}{50} + (15 - 12,2)^2 \frac{1}{50} = 1,68 \end{aligned}$$

3.4.1 Propiedades de la varianza

Se presentarán ahora varios resultados relacionados con las propiedades básicas de la varianza. En estos teoremas se supone que existe la varianza de todas las variables aleatorias. El primer teorema demuestra que la varianza de una variable aleatoria X no puede ser 0 a menos que la distribución de probabilidad de X se concentre en un sólo punto.

Teorema 3.1:

$Var(X) = 0$ si, y sólo si, existe una constante c tal que $P(X = c) = 1$.

Demostración:

Supóngase en primer lugar que existe una constante c tal que $P(X = c) = 1$. Entonces, $E(X) = c$ y $P[(X - c)^2 = 0] = 1$. Por tanto,

$$Var(X) = E[(X - c)^2] = 0.$$

Igualmente, supóngase que $Var(X) = 0$. Entonces, $P[(X - \mu)^2 \geq 0] = 1$ pero $E[(X - \mu)^2] = 0$. Por tanto, de acuerdo con el comentario que sigue a la demostración de la propiedad 2 de las propiedades de la esperanza, se puede observar que

$$P[(X - \mu)^2 = 0] = 1.$$

Por tanto

$$P(X = \mu) = 1$$

Teorema 3.2:

Para constantes a y b cualesquiera,

$$Var(aX + b) = a^2 Var(X)$$

Demostración:

Si $E(X) = \mu$, entonces $E(aX + b) = a\mu + b$. Por tanto,

$$\begin{aligned} Var(aX + b) &= E[(aX + b - a\mu - b)^2] = E[(aX - a\mu)^2] \\ &= a^2 E[(X - \mu)^2] = a^2 Var(X). \end{aligned}$$

Por el teorema anterior resulta que

$$Var(X + b) = Var(X)$$

para cualquier constante b . Este resultado es intuitivamente plausible, puesto que desplazando la distribución de X una distancia b unidades a lo largo de la recta real, la media de la distribución cambiará en b unidades, pero el desplazamiento no afectará a la dispersión de la distribución alrededor de su media.

Análogamente, del teorema anterior resulta que

$$\text{Var}(-X) = \text{Var}(X).$$

Este resultado también es intuitivamente plausible, puesto que reflejando la distribución de X con respecto al origen de la recta real, resultará una nueva distribución que será la imagen especular de la original. La media cambiará de μ a $-\mu$, pero la dispersión total alrededor de su media no se modificará.

El siguiente teorema proporciona un método alternativo para calcular el valor de $\text{Var}(X)$.

Teorema 3.3:

Para cualquier variable aleatoria X ,

$$\text{Var}(X) = E(X^2) - E^2(X)$$

Demostración:

Sea $E(X) = \mu$. Entonces,

$$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] = E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2 = E(X^2) - E^2(X). \end{aligned}$$

Ejemplo 3.13:

Supóngase que una variable aleatoria X puede tomar cada uno de los cinco valores $-2, 0, 1, 3$ y 4 con la misma probabilidad. Se determinará la desviación típica de $Y = 4X - 7$.

En este ejemplo,

$$E(X) = \frac{1}{5}(-2 + 0 + 1 + 3 + 4) = 1,2,$$

$$E(X^2) = \frac{1}{5}(4 + 0 + 1 + 9 + 16) = 6.$$

Entonces

$$\text{Var}(X) = E(X^2) - E^2(X) = 6 - (1,2)^2 = 4,56.$$

Y,

$$\text{Var}(Y) = 16\text{Var}(X) = 72,96.$$

Y por tanto la desviación típica, σ , de Y es

$$\sigma_Y = \sqrt{72,96} = 8,54.$$

3.5 Momentos

Para cualquier variable aleatoria X y cualquier entero positivo k , la esperanza $E(X^k)$ se denomina momento de orden k respecto al origen de X . En particular, de acuerdo con esta terminología, la media de X es el momento de orden uno de X . Se dice que el momento de orden k existe si, y sólo si, $E(|X|^k) < \infty$. Si la variable aleatoria X está acotada, esto es, si existen números finitos a y b tales que $P(a \leq X \leq b) = 1$, entonces deben existir necesariamente todos los momentos de X . Es posible, sin embargo, que existan todos los momentos de X aunque X no esté acotada.

3.5.1 Función generatriz de momentos

Considérese ahora una variable aleatoria X ; y para cada número real t se define

$$\psi(t) = E(e^{tX}) \quad (3.6)$$

La función ψ se denomina función generatriz de momentos de X . Si la variable aleatoria X está acotada, entonces para cualquier valor de t debe existir la esperanza de la ecuación 3,6. En este caso, por tanto, la función generatriz de momentos de la variable aleatoria X existirá para todos los valores de t . Por otro lado, si X no está acotada, entonces la función generatriz de momentos puede existir para algunos valores de t y no existir para otros. De la ecuación 3,6 se puede observar, sin embargo, que para cualquier variable aleatoria X la función generatriz de momentos $\psi(t)$ debe existir en el punto $t = 0$ y en ese punto su valor debe ser $\psi(0) = E(1) = 1$.

Ejemplo 3.14:

Sea X una variable aleatoria con función de probabilidad Bernoulli, es decir, $X \sim B(p)$, entonces, su función generatriz de momentos es:

$$\begin{aligned} \varphi_X(t) &= E[e^{tx}] = \sum_{i \in I} e^{ti} P(X = i) = e^{t1} p + e^{t0} (1 - p) \\ &= e^t p + (1 - p). \end{aligned}$$

Ejemplo 3.15:

Sea X una variable aleatoria con función de probabilidad Binomial, es decir, $X \sim B(n, p)$, entonces, su función generatriz de momentos es:

$$\begin{aligned} \varphi_X(t) &= E[e^{tx}] = \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= [pe^t + (1-p)]^n. \end{aligned}$$

Ejemplo 3.16:

Sea X una variable aleatoria con función de probabilidad Poisson, es decir, $X \sim P(\lambda)$, entonces, su función generatriz de momentos es:

$$\varphi_X(t) = E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} = e^{\lambda(e^t - 1)},$$

usando el desarrollo de Taylor¹ de la función e^x .

$$\begin{aligned} EX &= \frac{\phi'_X(0)}{i} = \lambda \\ EX^2 &= \frac{\phi''_X(0)}{i^2} = \lambda^2 + \lambda \\ \text{Var}(X) &= \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned}$$

Ejemplo 3.17:

Sea X una variable aleatoria con función de probabilidad Gamma, es decir, $X \sim \Gamma(a, p)$, entonces, su función generatriz de momentos es:

$$\phi_X(t) = \int_0^\infty e^{tx} \frac{p^a}{\Gamma(a)} e^{-px} X^{a-1} dx = \left(\frac{p}{p-t}\right)^a.$$

Supóngase que existe la función generatriz de momentos de una variable aleatoria X para todos los valores de t en un intervalo alrededor del punto $t = 0$. Entonces se puede demostrar que existe la derivada $\psi'(t)$ en el punto $t = 0$ y que en ese punto la derivada de la esperanza de la ecuación 3.6 debe ser igual a la esperanza de la derivada. Entonces,

$$\psi'(0) = \left[\frac{d}{dt} E(e^{tX})\right]_{t=0} = E\left[\left(\frac{d}{dt} e^{tX}\right)_{t=0}\right].$$

Pero, puesto que

$$\left(\frac{d}{dt} e^{tX}\right)_{t=0} = (X e^{tX})_{t=0} = X,$$

resulta que

$$\psi'(0) = E(X).$$

En otras palabras, la derivada de la función generatriz de momentos $\psi(t)$ en el punto $t = 0$ es la media de X .

En general, si la función generatriz de momentos de X existe para todos los valores de t en un intervalo alrededor del punto $t = 0$, entonces se puede demostrar que deben existir todos los momentos $E(X^k)$ de X ($k = 1, 2, \dots$). Además, se puede demostrar que es posible derivar $\psi(t)$ un número arbitrario de veces en el punto $t = 0$. Para $n = 1, 2, \dots$, la n -ésima derivada $\psi^{(n)}(0)$ en el punto $t = 0$ satisfará la relación siguiente:

$$\begin{aligned} \psi^{(n)}(0) &= \left[\frac{d^n}{dt^n} E(e^{tX})\right]_{t=0} = E\left[\left(\frac{d^n}{dt^n} e^{tX}\right)_{t=0}\right] \\ &= E[(X^n e^{tX})_{t=0}] = E(X^n). \end{aligned}$$

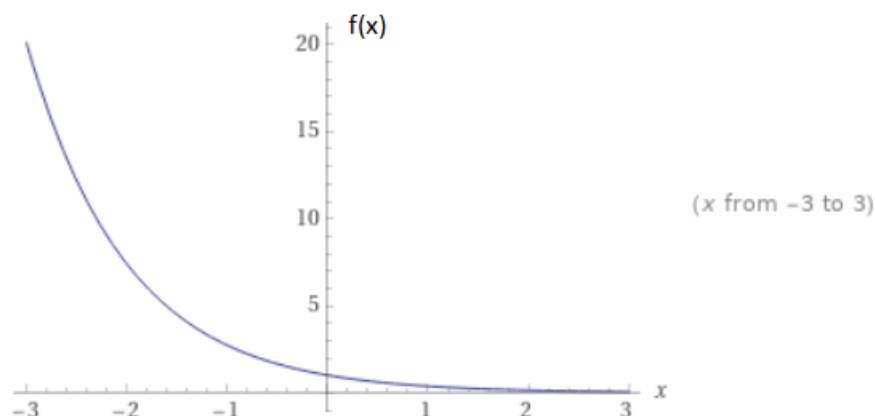
Entonces, $\psi'(0) = E(X)$, $\psi''(0) = E(X^2)$, $\psi'''(0) = E(X^3)$, y así sucesivamente.

¹ $\sum_{k=0}^\infty \frac{x^k}{k!} = e^x$ Desarrollo de Taylor de e^x

Ejemplo 3.18:

Supóngase que X es una variable aleatoria cuya función de densidad es la siguiente:

$$f(x) = \begin{cases} e^{-x} & \text{si } x > 0 \\ 0 & \text{fuera.} \end{cases}$$



Calcular la función generatriz de momentos y usarla para calcular la $\text{Var}(X)$.

Para cualquier número real t ,

$$\psi(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} e^{-x} dx = \int_0^{\infty} e^{(t-1)x} dx.$$

La última integral de esta ecuación será finita si, y sólo si, $t < 1$. Por tanto, $\psi(t)$ existe sólo para $t < 1$. Para tales valores de t ,

$$\psi(t) = \frac{1}{1-t}.$$

Puesto que $\psi(t)$ es finita para todos los valores de t en un intervalo alrededor del punto $t = 0$, existen todos los momentos de X . Las dos primeras derivadas de ψ son

$$\psi'(t) = \frac{1}{(1-t)^2}$$

y

$$\psi''(t) = \frac{2}{(1-t)^3}.$$

Por tanto, $E(X) = \psi'(0) = 1$ y $E(X^2) = \psi''(0) = 2$. Resulta ahora que

$$\text{Var}(X) = \psi''(0) - [\psi'(0)]^2 = 1$$

Propiedades de las funciones generatrices de momentos

Se presentarán ahora dos teoremas básicos relacionados con las funciones generatrices de momentos.

Teorema 3.4:

Sea X una variable aleatoria cuya función generatriz de momentos es ψ_1 ; y sea $Y = aX + b$, donde a y b son constantes; y sea ψ_2 la función generatriz de momentos de Y . Entonces, para cualquier valor de t tal que existe $\psi_1(at)$,

$$\psi_2(t) = e^{bt} \psi_1(at).$$

Demostración:

Por la definición de una función generatriz de momentos

$$\psi_2(t) = E(e^{tY}) = E[e^{t(aX+b)}] = e^{bt} E(e^{atX}) = e^{bt} \psi_1(at).$$

Ejemplo 3.19:

Supóngase que la distribución de X es la del ejemplo anterior. Entonces la función generatriz de momentos de X para $t < 1$ es

$$\psi_1(t) = \frac{1}{1-t}.$$

Si $Y = 3 - 2X$, entonces la función generatriz de Y existirá para $t > -\frac{1}{2}$ y tendrá el valor

$$\psi_2(t) = e^{3t} \psi_1(-2t) = \frac{e^{3t}}{1+2t}$$

Se enunciará ahora otra importante propiedad de la función generatriz de momentos. La demostración se omite por estar fuera del alcance de este temario.

Teorema 3.5:

Si las funciones generatrices de momentos de dos variables X_1 y X_2 son idénticas para todos los valores de t en un intervalo alrededor del punto $t = 0$, entonces las distribuciones de probabilidad de X_1 y X_2 deben ser idénticas. Esta propiedad nos dice que si existe la función generatriz de momentos, entonces caracteriza a la distribución

3.6 Función Característica

Es una herramienta muy importante para obtener momentos de las variables. También para probar la reproductividad de las distribuciones e incluso para analizar propiedades de la convergencia de sucesiones de variables aleatorias.

Definición 3.3:

Dada una variable aleatoria X se define su función característica $\phi_X(t)$ como la función compleja de variable real dada por:

$$\begin{aligned} \phi_X : \mathbb{R} &\rightarrow \mathbb{C} \\ t &\rightarrow \phi_X(t) = E(e^{itX}), \end{aligned}$$

donde $i = \sqrt{-1}$ y $E(e^{itX}) = \begin{cases} \sum e^{itj}P(X=j) & \text{si } X \text{ es discreta} \\ \int e^{itx}f(x)dx & \text{si } X \text{ es continua.} \end{cases}$

Propiedades:

a) Existe $\forall t \in \mathbb{R}$ y toda variable aleatoria X .

Demostración:

$$|\varphi_X(t)| = |E(e^{itX})| = \left| \int_{\mathbb{R}} e^{itx} f(x) dx \right| \leq \int_{\mathbb{R}} |e^{itx}| f(x) dx = \int_{\mathbb{R}} 1 f(x) dx = 1.$$

b) Es característica. No puede existir dos distribuciones distintas con la misma función característica. Dicho de otra manera, conocida la función característica de una variable, queda identificada de forma unívoca la distribución de la variable.

c) La función característica es función generatriz de momentos:

$$EX^n = \frac{\varphi_X^{(n)}(0)}{i^n}.$$

Ejemplo 3.20:

Sean X el número de caras al lanzar tres monedas. Obtener utilizando la función característica $E(X)$ y la $Var(X)$.

$$\varphi_X(t) = \sum e^{itj}P(X=j) = \frac{1}{8}(1 + 3e^{it} + 3e^{2it} + e^{3it})$$

$$\varphi_X'(t) = \frac{1}{8}(3ie^{it} + 6ie^{2it} + 3ie^{3it})$$

$$\varphi_X'(0) = \frac{1}{8}(3i + 6i + 3i) = \frac{3}{2}i$$

$$\varphi_X''(t) = \frac{1}{8}(3i^2e^{it} + 12i^2e^{2it} + 9i^2e^{3it})$$

$$\varphi_X''(0) = \frac{1}{8}(3i^2 + 12i^2 + 9i^2) = 3i^2$$

$$EX = \frac{\varphi_X'(0)}{i} = \frac{\frac{3}{2}i}{i} = \frac{3}{2}$$

$$EX^2 = \frac{\varphi_X''(0)}{i^2} = \frac{3i^2}{i^2} = 3$$

$$Var(X) = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}$$

Ejemplo 3.21:

Sea X una variable aleatoria con función de probabilidad Bernoulli, es decir,

$$X = \begin{cases} 0 & q = 1 - p \\ 1 & p. \end{cases}$$

Su función característica es:

$$\varphi_X(t) = e^{it0}(1-p) + e^{it1}p = q + pe^{it} \text{ (con } q=1-p \text{)}.$$

Calculamos su media y varianza a partir de la función característica.

$$EX = \frac{\varphi'_X(0)}{i} = \frac{p \cdot i}{i} = p$$

$$EX^2 = \frac{\varphi''_X(0)}{i^2} = \frac{p \cdot i^2}{i^2} = p$$

$$\text{Var}(X) = p - p^2 = pq$$

Ejemplo 3.22:

Sea X una variable aleatoria con función de probabilidad Binomial, es decir,

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, 2, \dots, n.$$

Su función característica es:

$$\varphi_X(t) = E(e^{itX}) = \sum_{k=0}^n e^{itk} \binom{n}{k} p^k (1-p)^{n-k} = [pe^{it} + (1-p)]^n.$$

Calculamos su media y varianza a partir de la función característica.

$$EX = \frac{\varphi'_X(0)}{i} = np$$

$$EX^2 = \frac{\varphi''_X(0)}{i^2} = n(n-1)p^2 + np$$

$$\text{Var}(X) = n(n-1)p^2 + np - (np)^2 = npq.$$

Ejemplo 3.23:

Sea X una variable aleatoria con función de probabilidad Poisson, es decir,

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, 2, \dots$$

Su función característica es: $\varphi_X(t) = E(e^{itX}) = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} = e^{\lambda(e^{it}-1)}$, usando el desarrollo de Taylor² de la función e^x .

Calculamos su media y varianza a partir de la función característica.

$$EX = \frac{\varphi'_X(0)}{i} = \lambda$$

² $\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$ Desarrollo de Taylor de e^x

$$EX^2 = \frac{\phi_X''(0)}{i^2} = \lambda^2 + \lambda$$

$$\text{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Ejemplo 3.24:

Sea X una variable aleatoria con función de probabilidad Gamma, es decir, $X \sim \Gamma(a, p)$, entonces, su función generatriz de momentos es:

$$f(x) = \begin{cases} \frac{p^a}{\Gamma(a)} e^{-px} x^{a-1} & \text{si } x, a, p > 0 \\ 0 & \text{fuera,} \end{cases}$$

donde $\Gamma(a) = \int_0^\infty e^{-x} x^{a-1} dx = (a-1)!$ si $a \in \mathbb{N}$

Su función característica es:

$$\phi_X(t) = \int_0^\infty e^{itx} \frac{p^a}{\Gamma(a)} e^{-px} x^{a-1} dx = \left(\frac{p}{p - it} \right)^a$$

Calculamos su media y varianza a partir de la función característica.

$$EX = \frac{\phi_X'(0)}{i} = \frac{a}{p}$$

$$EX^2 = \frac{\phi_X''(0)}{i^2} = \frac{a(a+1)}{p^2}$$

$$\text{Var}(X) = \frac{a(a+1)}{p^2} - (a/p)^2 = \frac{a}{p^2}$$

Ejemplo 3.25:

Puede determinarse que si $X \sim N(\mu, \sigma)$ su función característica es

$$\phi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}.$$

Los cálculos son tediosos, aunque no difíciles, y serán omitidos aquí.

3.7 Desigualdades de Markov y Chebyshev

En esta sección se presentan dos resultados sencillos y generales, conocidos como la desigualdad de Markov y la desigualdad de Chebyshev.

3.7.1 Desigualdad de Markov

Supóngase que X es una variable aleatoria tal que $P(X \geq 0) = 1$. Entonces para cualquier número $t > 0$,

$$P(X \geq t) \leq \frac{E(X)}{t}.$$

Demostración:

Por conveniencia, supóngase que X tiene una distribución discreta cuya función de probabilidad

es $P(X = i)$. La demostración para una distribución continua o un tipo de distribución más general es análoga. Entonces,

$$E(X) = \sum_{i \in I} iP(X = i) = \sum_{i < t} iP(X = i) + \sum_{i \geq t} iP(X = i),$$

verifica para cualquier distribución.

Puesto que X puede tomar únicamente valores no negativos, todos los términos de la suma son no negativos. Por tanto,

$$E(X) = \sum_{i \in I} iP(X = i) \geq \sum_{i \geq t} tP(X = i) = tP(X \geq t).$$

La desigualdad de Markov tiene fundamentalmente interés para valores grandes de t . De hecho, cuando $t \leq E(X)$, la desigualdad no tiene interés alguno, puesto que se sabe que $P(X \geq t) \leq 1$. Sin embargo, de la desigualdad de Markov se obtiene que para cualquier variable aleatoria no negativa X cuya media es 1, el valor máximo posible de $P(X \geq 100)$ es 0,01. Además, se puede verificar que este valor máximo es alcanzado por la variable X para la cual $P(X = 0) = 0,99$ y $P(X = 100) = 0,01$.

3.7.2 Desigualdad de Chebyshev:

Sea X una variable aleatoria para la cual $Var(X)$ existe. Entonces para cualquier número concreto $t > 0$,

$$P(|X - EX| \geq t) \leq \frac{Var(X)}{t^2}$$

Demostración:

Sea $Y = [X - E(X)]^2$. Entonces $P(Y \geq 0) = 1$ y $E(Y) = Var(X)$. Aplicando a Y la desigualdad de Markov, se obtiene el siguiente resultado:

$$P(|X - EX| \geq t) = P(Y \geq t^2) \leq \frac{EY}{t^2} = \frac{Var(X)}{t^2}.$$

Se puede ver en la demostración que la desigualdad de Chebyshev es simplemente un caso especial de la desigualdad de Markov. Por tanto, los comentarios que siguen a la demostración de la desigualdad de Markov se pueden aplicar también a la desigualdad de Chebyshev. Estas desigualdades son muy útiles debido a su generalidad. Por ejemplo, si $Var(X) = \sigma^2$ y se define $t = 3\sigma$, entonces la desigualdad de Chebyshev proporciona el resultado

$$P(|X - EX| \geq 3\sigma) \leq \frac{1}{9}.$$

En otras palabras, este resultado afirma que la probabilidad de que cualquier variable aleatoria difiera de su media en más de 3 desviaciones estándar no puede exceder de $\frac{1}{9}$. Esta probabilidad será realmente mucho menor que $\frac{1}{9}$ para muchas de las variables aleatorias y distribuciones.

En la siguiente tabla se muestra la cantidad máxima y mínima de porcentaje de fuera y dentro de una cantidad fija $t = a$ de desviaciones estándar a la media.

| a | Mín. % dentro de a desviaciones estándar de la media | Máx. % fuera de a desviaciones estándar de la media |
|------------|--|---|
| 1 | 0% | 100% |
| $\sqrt{2}$ | 50% | 50% |
| 1.5 | 55.56% | 44.44% |
| 2 | 75% | 25% |
| 3 | 88.8889% | 11.1111% |
| 4 | 93.75% | 6.25% |
| 5 | 96% | 4% |
| 6 | 97.2222% | 2.7778% |
| 7 | 97.9592% | 2.0408% |
| 8 | 98.4375% | 1.5625% |
| 9 | 98.7654% | 1.2346% |
| 10 | 99% | 1% |

Ejemplo 3.26:

Para ilustrar este resultado, supongamos que los artículos de Wikipedia tienen una extensión media de 1000 caracteres y una desviación típica de 200 caracteres. De la desigualdad de Tchebychev, usando $a = 2$, se deduce que al menos el 75 % de los artículos tendrán una extensión comprendida entre 600 y 1400 caracteres.

Ejemplo 3.27:

Otra consecuencia de la desigualdad de Chebyshev es que para cada distribución de media y desviación típica finita, al menos la mitad de sus valores se concentrarán en el intervalo

$$(\mu - \sqrt{2}\sigma, \mu + \sqrt{2}\sigma),$$

ya que

$$P(|X - \mu| > \sqrt{2}\sigma) \leq \frac{\sigma^2}{2\sigma^2} = \frac{1}{2},$$

y lógicamente

$$P(|X - \mu| \leq \sqrt{2}\sigma) \geq \frac{1}{2}.$$

Ejemplo 3.28:

Si aplicamos la desigualdad a la distribución $N(\mu, \sigma)$ obtenemos

$$P\left(\left|\frac{X - \mu}{\sigma}\right| > 1,96\right) = P(|Z| > 1,96) \leq \frac{1}{1,96^2} = 0,26.$$

Si se compara este valor con el real $P(|Z| > 1,96) = 0,05$ se puede apreciar que, como se ha dicho anteriormente, la cota puede resultar muy burda.

Bibliografía

1. A.A. Borovkov (2013). *Probability Theory*. Springer.
2. A. Klenke (2014). *Probability Theory. A Comprehensive Course*. Springer.
3. M.H. DeGroot (1988). *Probability and Statistics*. Wilmington, Delaware, E.U.A. Addison-Wesley Iberoamericana, S.A.
4. V.K. Rohatgi and A.K. Md. Ehsanes Saleh (2001). *An Introduction to Probability and Statistics*. John Wiley and Sons.

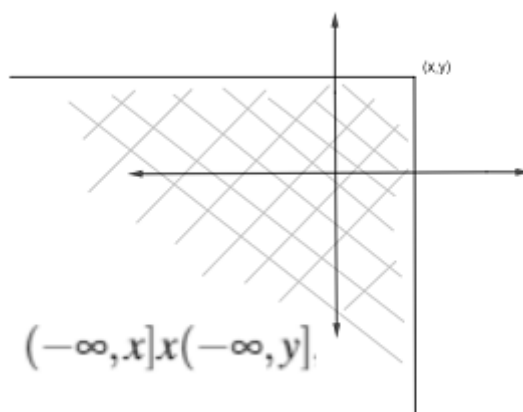
Tema 4

Distribuciones de varias variables aleatorias. Distribuciones conjuntas y marginales. Independencia entre variables aleatorias. Ejemplos

Lógicamente, con frecuencia en estadística es más interesante disponer de varias variables y analizar la posible vinculación entre ellas. Este tema está dedicado a ese objetivo.

Definición 4.1:

La σ -álgebra de Borel en \mathbb{R}^2 es la familia de subconjuntos de \mathbb{R}^2 que se pueden generar a partir de los conjuntos de la forma $(-\infty, x] \times (-\infty, y]$, cuando $x, y \in \mathbb{R}$.



Cuando nuestro espacio muestral Ω sea finito o contable utilizaremos como σ -álgebra 2^Ω , cuando sea \mathbb{R} utilizaremos la σ -álgebra de Borel y cuando sea \mathbb{R}^2 utilizaremos la σ -álgebra de Borel en \mathbb{R}^2 .

Definición 4.2:

Dado un espacio probabilístico $(\Omega, \sigma(\Omega), P)$ llamaremos vector aleatorio bidimensional, notado (X, Y) a cualquier función:

$$(X, Y) : \Omega \longrightarrow \mathbb{R}^2$$
$$\omega \in \Omega \longrightarrow (X(\omega), Y(\omega)),$$

que satisfaga $\forall x, y \in \mathbb{R}, \{\omega | X(\omega) \leq x, Y(\omega) \leq y\} = \{X \leq x, Y \leq y\} \in \sigma(\Omega)$.

Ejemplo 4.1:

Consideramos el experimento aleatorio consistente en lanzar dos dados equilibrados de forma

independiente. Sea (X, Y) el vector suma y diferencia (con signo) de los puntos obtenidos.

$$\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (6, 6)\}, |\Omega| = 6^2 = 36$$

$$P : 2^\Omega \longrightarrow [0, 1]$$

$$A \longrightarrow P(A) = \frac{\text{casos favorables de } A}{\text{casos posibles}} = \frac{|A|}{36}, \text{ (Regla de Laplace).}$$

$$(X, Y) : (\Omega, \sigma(\Omega), P) \longrightarrow \mathbb{R}^2$$

$$(i, j) \longrightarrow (i + j, i - j), i, j \in \{1, 2, 3, 4, 5, 6\}$$

i) Si $x = 4, y = -4$, $\{\omega | X(\omega) \leq 4, Y(\omega) \leq -4\} = \emptyset \in \sigma(\Omega)$.

ii) Si $x = 5, y = 1$, $\{\omega | X(\omega) \leq 5, Y(\omega) \leq 1\} = \{(1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (2, 2), (2, 3), (3, 2)\} \in \sigma(\Omega)$.

Como $\sigma(\Omega)$ es 2^Ω entonces en $\sigma(\Omega)$ estarán todos los subconjuntos de Ω , y por tanto, $\forall x, y, \{\omega | X \leq x, Y \leq y\} \in \sigma(\Omega)$. Luego (X, Y) es un vector aleatorio bidimensional.

Proposición 4.1:

Sea $(\Omega, \sigma(\Omega), P)$ un espacio probabilístico y sea (X, Y) un vector aleatorio bidimensional definido en Ω . Entonces X e Y son variables aleatorias.

Demostración: En un espacio probabilístico $(\Omega, \sigma(\Omega), P)$, sabemos que X es una variable aleatoria si $\forall x \in \mathbb{R}, \{\omega | X(\omega) \leq x\} \in \sigma(\Omega)$. En particular

$$\{\omega | X(\omega) \leq x\} = \{\omega | X(\omega) \leq x, Y(\omega) < \infty\} \in \sigma(\Omega),$$

por definición de variable aleatoria bidimensional (X, Y) . Simétricamente

$$\{\omega | Y(\omega) \leq y\} = \{\omega | Y(\omega) \leq y, X(\omega) < \infty\} \in \sigma(\Omega),$$

En consecuencia X e Y son variables aleatorias. Por tanto, hablaremos indistintamente de vector aleatorio bidimensional o de dos variables aleatorias. Cada una de las variables que forma un vector bidimensional recibe el nombre de variable marginal.

4.1 Función de distribución de un vector aleatorio bidimensional.

Definición 4.3:

Dado un espacio probabilístico $(\Omega, \sigma(\Omega), P)$ y un vector aleatorio (X, Y) , llamaremos función de distribución de (X, Y) a la función

$$F : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$(x, y) \rightarrow P(X \leq x, Y \leq y).$$

Ejemplo 4.2:

En el ejemplo anterior (suma y diferencia de puntos del lanzamiento de dos dados) tenemos que:

$$F(4, -4) = P(X \leq 4, Y \leq -4) = P(\emptyset) = 0.$$

$$F(5, 1) = P(X \leq 5, Y \leq 1) = P(\{(1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (2, 2), (2, 3), (3, 2)\}) = \frac{8}{36} = \frac{2}{9}.$$

4.1.1 Propiedades de la función de distribución.

Sea (X, Y) un vector aleatorio bidimensional definido en un espacio probabilístico $(\Omega, \sigma(\Omega), P)$ y F su función de distribución conjunta. Entonces:

$$i) \forall x, y \in \mathbb{R}, F(x, y) \in [0, 1].$$

Demostración:

$\forall x, y \in \mathbb{R}, F(x, y) = P(X \leq x, Y \leq y) \in [0, 1]$ ya que es una probabilidad.

ii) Es no decreciente en cada variable, es decir:

$$\text{Si } x \leq x_0, F(x, y) \leq F(x_0, y), \forall y \text{ y si } y \leq y_0, F(x, y) \leq F(x, y_0), \forall x.$$

Demostración:

Si $x \leq x_0$, $F(x, y) = P(\{\omega | X(\omega) \leq x, Y(\omega) \leq y\})$ y $F(x_0, y) = P(\{\omega | X(\omega) \leq x_0, Y(\omega) \leq y\})$. Ahora bien, como $x \leq x_0$, $\{\omega | X(\omega) \leq x, Y(\omega) \leq y\} \subseteq \{\omega | X(\omega) \leq x_0, Y(\omega) \leq y\}$ por lo que se obtiene que $P(X \leq x, Y \leq y) \leq P(X \leq x_0, Y \leq y)$, es decir, $F(x, y) \leq F(x_0, y)$.

iii)

$$\lim_{x \rightarrow \infty} F(x, y) = \lim_{x \rightarrow \infty} P(X \leq x, Y \leq y) = P(Y \leq y) = F(y),$$

$$\lim_{y \rightarrow \infty} F(x, y) = \lim_{y \rightarrow \infty} P(X \leq x, Y \leq y) = P(X \leq x) = F(x).$$

iv)

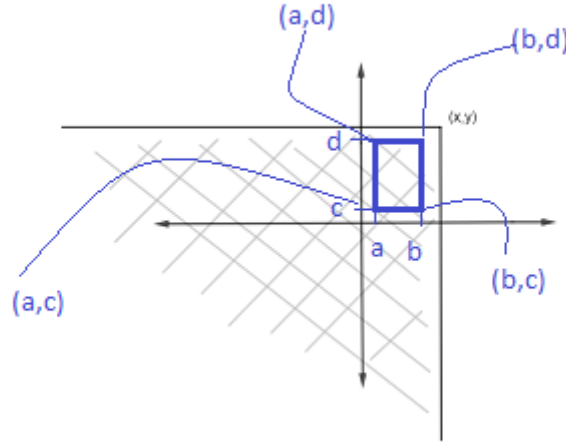
$$\lim_{x \rightarrow -\infty} F(x, y) = \lim_{x \rightarrow -\infty} P(X \leq x, Y \leq y) = P(X \leq -\infty, Y \leq y) = P(\emptyset) = 0,$$

$$\lim_{y \rightarrow -\infty} F(x, y) = \lim_{y \rightarrow -\infty} P(X \leq x, Y \leq y) = P(X \leq x, Y \leq -\infty) = P(\emptyset) = 0.$$

v)

$$\lim_{x \rightarrow \infty} \lim_{y \rightarrow \infty} F(x, y) = \lim_{y \rightarrow \infty} \lim_{x \rightarrow \infty} F(x, y) = 1$$

vi) Cálculo de probabilidades de rectángulos utilizando la función de distribución. Si $a, b, c, d \in \mathbb{R}$, $P(a \leq X \leq b, c \leq Y \leq d)$, se calcula como:



$$P(a \leq X \leq b, c \leq Y \leq d) = P(X \leq b, Y \leq d) - P(X \leq a, Y \leq d) - P(X \leq b, Y \leq c) + P(X \leq a, Y \leq c) = F(b, d) - F(a, d) - F(b, c) + F(a, c).$$

Vectores aleatorios bidimensionales discretos. Función de probabilidad conjunta y funciones de probabilidad marginales

Definición 4.4:

Diremos que el vector aleatorio (X, Y) es discreto si sus variables marginales X e Y son discretas, es decir, cada una de ellas toma una cantidad finita o numerable de valores.

Para vectores aleatorios discretos se define la llamada función de probabilidad (masa o cuantía) conjunta como:

$$P_{i,j} = P(X = i, Y = j), i \in I, j \in J,$$

siendo I el conjunto de valores que tomará X y J el conjunto de valores que tomará Y . Dicha función satisface:

$$i) P_{i,j} = P(X = i, Y = j) \geq 0, \forall i \in I, \forall j \in J.$$

$$ii) \sum_{i \in I} \sum_{j \in J} P_{i,j} = \sum_{i \in I} \sum_{j \in J} P(X = i, Y = j) = 1.$$

Ambas condiciones se deducen trivialmente de la definición de probabilidad ya que:

$$i) P_{i,j} = P(X = i, Y = j) \geq 0, \text{ dado que es una probabilidad.}$$

$$ii) \sum_{i \in I} \sum_{j \in J} P(X = i, Y = j) = P(\Omega) = 1.$$

La función de probabilidad de la variable X (función de probabilidad marginal de X) es:

$$P(X = i) = \sum_{j \in J} P(X = i, Y = j) \text{ si } i \in I.$$

La función de probabilidad de la variable Y (función de probabilidad marginal de Y) será:

$$P(Y = j) = \sum_{i \in I} P(X = i, Y = j) \text{ si } j \in J.$$

Ejemplo 4.3:

Se lanzan dos dados. Sea X la suma de puntos obtenidos e Y la diferencia entre los puntos del primer dado y el segundo (con signo). Obtengamos la función de probabilidad conjunta y las funciones de probabilidad marginales.

Si las variables no toman demasiados valores se acostumbra a escribir la función de probabilidad conjunta en forma de tabla y las funciones de probabilidad marginales se sitúan en los márgenes de dicha tabla, de ahí su nombre.

| XY | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | X |
|----|------|------|------|------|------|------|------|------|------|------|------|------|
| 2 | 0 | 0 | 0 | 0 | 0 | 1/36 | 0 | 0 | 0 | 0 | 0 | 1/36 |
| 3 | 0 | 0 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 0 | 0 | 2/36 |
| 4 | 0 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 0 | 3/36 |
| 5 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 4/36 |
| 6 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 5/36 |
| 7 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 6/36 |
| 8 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 5/36 |
| 9 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 4/36 |
| 10 | 0 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 0 | 3/36 |
| 11 | 0 | 0 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 0 | 0 | 2/36 |
| 12 | 0 | 0 | 0 | 0 | 0 | 1/36 | 0 | 0 | 0 | 0 | 0 | 1/36 |
| Y | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 6/36 | 5/36 | 4/36 | 3/36 | 2/36 | 1/36 | 1 |

X e Y no están igualmente distribuidas.

Ejemplo 4.4:

Se lanzan dos dados. Sea X el resultado del primer dado e Y el máximo de los puntos obtenidos en ambos dados. Obtener la función de probabilidad conjunta y las marginales del vector (X, Y) .

| XY | 1 | 2 | 3 | 4 | 5 | 6 | X |
|----|------|------|------|------|------|-------|------|
| 1 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 6/36 |
| 2 | 0 | 2/36 | 1/36 | 1/36 | 1/36 | 1/36 | 6/36 |
| 3 | 0 | 0 | 3/36 | 1/36 | 1/36 | 1/36 | 6/36 |
| 4 | 0 | 0 | 0 | 4/36 | 1/36 | 1/36 | 6/36 |
| 5 | 0 | 0 | 0 | 0 | 5/36 | 1/36 | 6/36 |
| 6 | 0 | 0 | 0 | 0 | 0 | 6/36 | 6/36 |
| Y | 1/36 | 3/36 | 5/36 | 7/36 | 9/36 | 11/36 | 1 |

Ejemplo 4.5:

Se lanzan dos dados. Sea X el mínimo de los puntos obtenidos en ambos dados e Y el máximo. Obtener la función de probabilidad conjunta y las marginales del vector (X, Y) .

| XY | 1 | 2 | 3 | 4 | 5 | 6 | X |
|----|------|------|------|------|------|-------|-------|
| 1 | 1/36 | 2/36 | 2/36 | 2/36 | 2/36 | 2/36 | 11/36 |
| 2 | 0 | 1/36 | 2/36 | 2/36 | 2/36 | 2/36 | 9/36 |
| 3 | 0 | 0 | 1/36 | 2/36 | 2/36 | 2/36 | 7/36 |
| 4 | 0 | 0 | 0 | 1/36 | 2/36 | 2/36 | 5/36 |
| 5 | 0 | 0 | 0 | 0 | 1/36 | 2/36 | 3/36 |
| 6 | 0 | 0 | 0 | 0 | 0 | 1/36 | 1/36 |
| Y | 1/36 | 3/36 | 5/36 | 7/36 | 9/36 | 11/36 | 1 |

Ejemplo 4.6:

Se lanzan dos dados. Sea X el resultado del primer dado e Y la diferencia en módulo entre los resultados de ambos dados.

Calcular:

- a) $p(X + Y > 6)$
- b) $p(X + Y \geq 6)$
- c) $p(|X - Y| \leq 1)$

| XY | 0 | 1 | 2 | 3 | 4 | 5 | X |
|----|------|-------|------|------|------|------|------|
| 1 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 6/36 |
| 2 | 1/36 | 2/36 | 1/36 | 1/36 | 1/36 | 0 | 6/36 |
| 3 | 1/36 | 2/36 | 2/36 | 1/36 | 0 | 0 | 6/36 |
| 4 | 1/36 | 2/36 | 2/36 | 1/36 | 0 | 0 | 6/36 |
| 5 | 1/36 | 2/36 | 1/36 | 1/36 | 1/36 | 0 | 6/36 |
| 6 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 6/36 |
| Y | 6/36 | 10/36 | 8/36 | 6/36 | 4/36 | 2/36 | 1 |

$$a) P(X + Y > 6) = P(X = 4, Y = 3) + P(X = 5, Y = 2) + P(X = 5, Y = 3) + P(X = 5, Y = 4) + P(X = 6, Y = 1) + P(X = 6, Y = 2) + P(X = 6, Y = 3) + P(X = 6, Y = 4) + P(X = 6, Y = 5) = 9/36 = 1/4$$

$$b) P(X + Y \geq 6) = P(X + Y > 6) + P(X = 1, Y = 5) + P(X = 2, Y = 4) + P(X = 3, Y = 3) + P(X = 4, Y = 2) + P(X = 5, Y = 1) + P(X = 6, Y = 0) = 1/4 + 8/36 = 17/36$$

$$c) P(|X - Y| \leq 1) = P(X = 1, Y = 0) + P(X = 1, Y = 1) + P(X = 1, Y = 2) + \dots = 13/36$$

Vectores aleatorios absolutamente continuos. Función de densidad conjunta y funciones de densidad marginales

Definición 4.5:

Diremos que un vector aleatorio (X, Y) es absolutamente continuo cuando su función de distribución $F(x, y)$ es continua, es decir, $\forall (x_0, y_0) \in \mathbb{R}^2$, $\lim_{x \rightarrow x_0, y \rightarrow y_0} F(x, y) = F(x_0, y_0)$.

En tal caso existe una función $f(x, y)$, llamada función de densidad conjunta, tal que $\forall x, y \in \mathbb{R}$

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) ds dt \Leftrightarrow \frac{\partial^2}{\partial x \partial y} F(x, y) = f(x, y)$$

La función de densidad conjunta verifica las dos siguientes propiedades:

i) $f(x, y) \geq 0$, $\forall x, y \in \mathbb{R}$, salvo a los sumo en un conjunto numerable de puntos.

Demostración: Si existiera un conjunto A no numerable con $f(x, y) < 0$, $\forall x, y \in A$, entonces $P(A) = \iint_A f(x, y) dx dy < 0$, lo cual contradice la definición de probabilidad

ii) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$.

Demostración: Se verifica pues $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = \iint_{\mathbb{R}^2} f(x, y) dx dy = P(\Omega) = 1$

Dada la densidad conjunta $f(x, y)$, las densidades marginales se calculan como:

$$f(x) = \int_{-\infty}^{+\infty} f(x, y) dy.$$

$$f(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

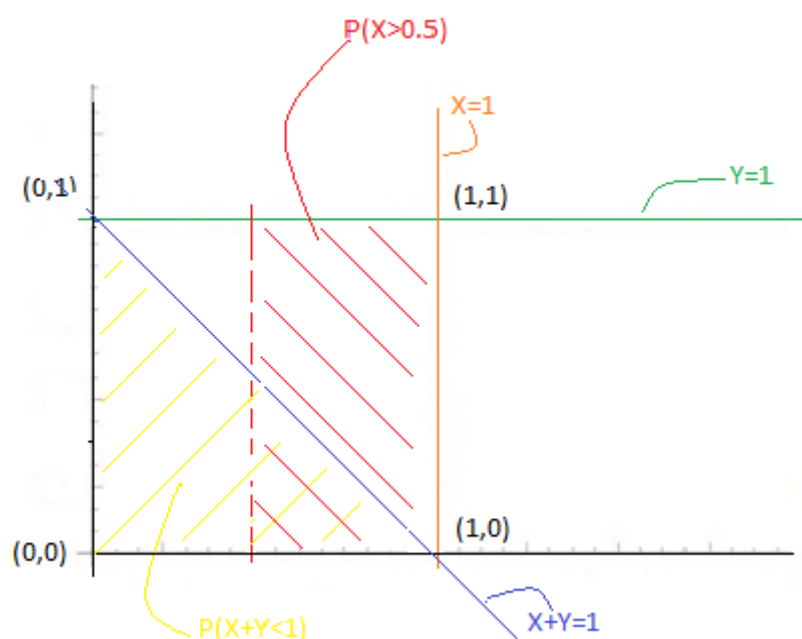
Ejemplo 4.7:

Sea (X, Y) un vector aleatorio bidimensional con densidad conjunta.

$$f(x, y) = \begin{cases} kx(1-y) & \text{si } x, y \in [0, 1] \\ 0 & \text{fuera.} \end{cases}$$

Determinar:

- El valor de k
- $P(X > 0,5)$
- $P(X + Y < 1)$
- Marginales



a)

$$\int \int_{\mathbb{R}^2} f(x,y) dx dy = 1 = \int_0^1 \int_0^1 kx(1-y) dx dy = 1; \text{ de donde } k = 4.$$

b)

$$P(X > 0,5) = \int_{0,5}^1 \int_0^1 4x(1-y) dy dx = 3/4.$$

c)

$$P(X + Y < 1) = \int_0^1 \int_0^{1-x} 4x(1-y) dy dx = 1/2.$$

Alternativamente:

$$P(X + Y < 1) = \int_0^1 \int_0^{1-y} 4x(1-y) dx dy = 1/2.$$

d) Marginales:

$$f(x) = \int_{-\infty}^{+\infty} f(x,y) dy = \int_0^1 f(x,y) dy = \begin{cases} 2x, & \text{si } x \in [0, 1] \\ 0, & \text{en otro caso.} \end{cases}$$

$$f(y) = \int_{-\infty}^{+\infty} f(x,y) dx = \int_0^1 f(x,y) dx = \begin{cases} 2y-2, & \text{si } y \in [0, 1] \\ 0, & \text{en otro caso.} \end{cases}$$

Otra forma de calcular el apartado b) teniendo la función de densidad marginal de X es:

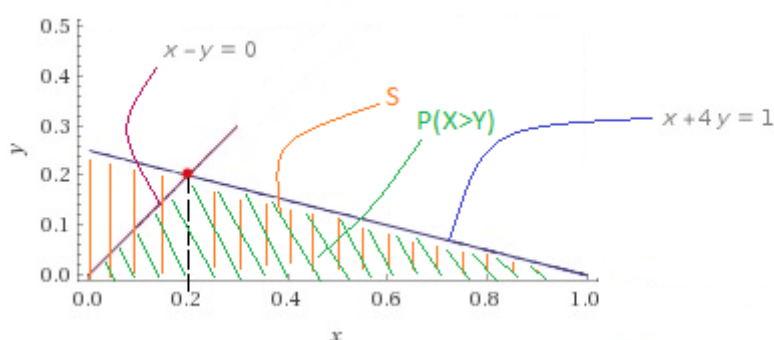
$$P(X > 0,5) = \int_{\frac{1}{2}}^1 2x dx = 3/4.$$

Ejemplo 4.8:

Se selecciona al azar un punto en $S = \{(x, y) \in \mathbb{R}^2 | x \geq 0, y \geq 0, 4y + x \leq 1\}$.

Calcular:

- a) La densidad $f(x, y)$
- b) $P(X > Y)$
- c) $P(X + Y > 1)$
- d) Marginales



a)

$$f(x, y) = \begin{cases} k & \text{si } (x, y) \in S \\ 0 & \text{fuera,} \end{cases}$$

ya que si la selección es al azar todos los puntos tienen que tener la misma densidad.

$$\iint_S f(x, y) dx dy = 1 = \iint_S k dx dy = k \cdot \text{Area}(S) = 1; \text{Area}(S) = \frac{1 \cdot \frac{1}{4}}{2} = \frac{1}{8}; k = 8.$$

b)

$$P(X > Y) = \int_0^{\frac{1}{5}} \int_0^x 8 dy dx + \int_{\frac{1}{5}}^1 \int_0^{\frac{1-x}{4}} 8 dy dx = \frac{4}{5}.$$

c) $P(X + Y > 1) = 0$.

d) Marginales:

$$f(x) = \int_{-\infty}^{+\infty} f(x,y) dy = \int_0^{\frac{1-x}{4}} 8 dy = \begin{cases} 2-2x & \text{si } x \in [0,1] \\ 0 & \text{fuera.} \end{cases}$$

$$f(y) = \int_{-\infty}^{+\infty} f(x,y) dx = \int_0^{1-4y} 8 dx = \begin{cases} 8(1-4y) & \text{si } y \in [0, \frac{1}{4}] \\ 0 & \text{fuera.} \end{cases}$$

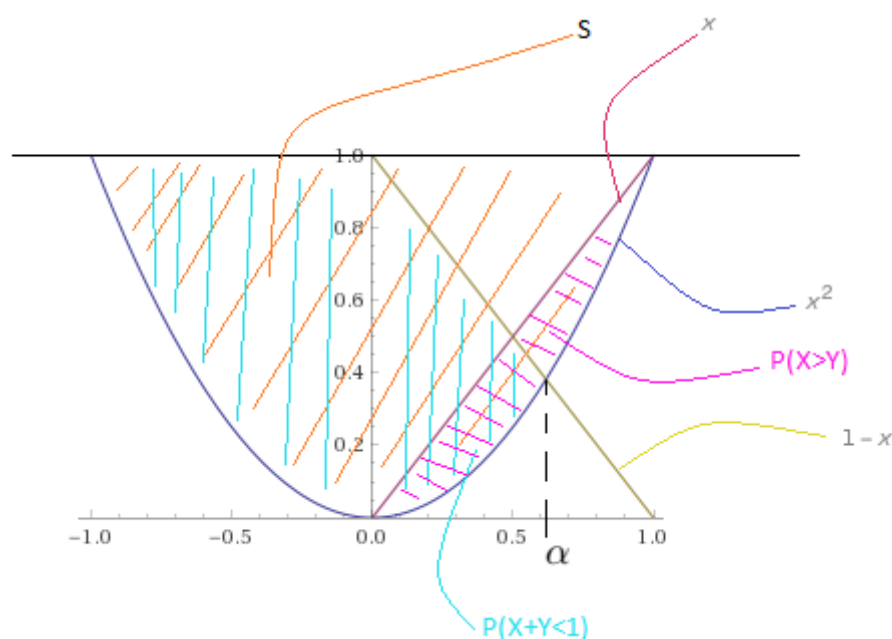
Ejemplo 4.9:

Sea (X,Y) un vector aleatorio bidimensional con densidad conjunta.

$$f(x,y) = \begin{cases} cx^2y & \text{si } x^2 \leq y \leq 1 \\ 0 & \text{fuera} \end{cases}$$

Calcular:

- El valor de c .
- $p(X > Y)$
- $p(X + Y < 1)$
- Marginales.



$$a) \iint_{\mathbb{R}^2} f(x,y) dx dy = 1 = \int_{-1}^1 \int_{x^2}^1 cx^2y dy dx = \frac{4c}{21}; \text{ de donde } c = \frac{21}{4}$$

b)

$$P(X > Y) = \int_0^1 \int_{x^2}^x \frac{21}{4} x^2 y dy dx = 3/20.$$

c) $P(X + Y < 1)$; Primero calculamos el valor del punto de corte entre la parábola $y = x^2$ y la recta $y = 1 - x$; a dicho punto lo notaremos como " α ",

$$1-x=x^2, x=\begin{cases} \frac{-1+\sqrt{5}}{2}=\alpha \\ \frac{-1-\sqrt{5}}{2}\notin S \end{cases}$$

$$P(X+Y < 1) = \int_{-1}^0 \int_{x^2}^1 \frac{21}{4}x^2y \, dydx + \int_0^\alpha \int_{x^2}^{1-x} \frac{21}{4}x^2y \, dydx = {}^1 \\ = \frac{1}{2} + \frac{21}{8} \left(\frac{\alpha^5}{5} - \frac{\alpha^4}{2} + \frac{\alpha^3}{3} - \frac{\alpha^7}{7} \right).$$

d) Marginales:

$$f(x) = \int_{-\infty}^{+\infty} f(x,y) \, dy = \int_{x^2}^1 \frac{21}{4}x^2y \, dy = \begin{cases} \frac{21}{8}x^2(1-x^4) & \text{si } x \in [-1, 1] \\ 0 & \text{fuera} \end{cases}$$

$$f(y) = \int_{-\infty}^{+\infty} f(x,y) \, dx = \int_{-\sqrt{y}}^{+\sqrt{y}} \frac{21}{4}x^2y \, dx = \begin{cases} \frac{7}{2}y^2\sqrt{y} & \text{si } y \in [0, 1] \\ 0 & \text{fuera} \end{cases}$$

Vectores aleatorios no discretos no absolutamente continuos

Existen vectores que no son discretos ni absolutamente continuos. En tal caso la distribución va a venir dada por una mixtura de una función de probabilidad y de una función de densidad. La probabilidad de los sucesos se obtendrá sumando las probabilidades de la parte discreta con la integral en la parte continua. Consideramos el siguiente ejemplo.

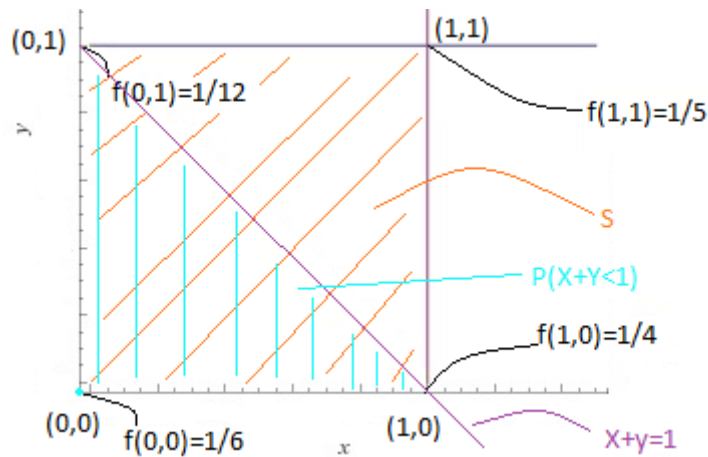
Ejemplo 4.10:

Se elige un punto en el cuadrado unitario $0 \leq x \leq 1$, $0 \leq y \leq 1$, la probabilidad de que sea $(0,0)$ es $1/6$, la de que sea $(0,1)$ es $1/12$, la de $(1,0)$ es $1/4$ y la de $(1,1)$ es $1/5$ y todos los demás puntos son igualmente verosímiles.

Sea (X,Y) el vector aleatorio que representa el punto seleccionado. (X,Y) no es discreto ya que el conjunto de valores que toma no es numerable. Por otro lado (X,Y) no es absolutamente continuo pues hay puntos que tienen probabilidad estrictamente positiva. En este caso $f(x,y)$, que no es función de densidad ni función de probabilidad, viene dada por:

$$f(x,y) = \begin{cases} 1/6 & \text{si } (x,y) = (0,0) \\ 1/4 & \text{si } (x,y) = (1,0) \\ 1/12 & \text{si } (x,y) = (0,1) \\ 1/8 & \text{si } (x,y) = (1,1) \\ k & \text{si } (x,y) \in S \\ 0 & \text{fuera} \end{cases}$$

¹ $\int_{-1}^0 \int_{x^2}^1 \frac{21}{4}x^2y \, dydx = \frac{1}{2}$ dada la simetría que presenta la función de densidad con respecto al eje Y



Para determinar k :

$$1 = \frac{1}{6} + \frac{1}{4} + \frac{1}{8} + \frac{1}{12} + \int_0^1 \int_0^{1-x} k \, dy \, dx = \frac{4+6+3+2}{24} + k \text{Area}(S); \text{ Por tanto, } k = 3/8.$$

$$P(X+Y < 1) = f(0,0) + \int_0^1 \int_0^{1-x} \frac{3}{8} k \, dy \, dx = \frac{17}{48}.$$

$$P(X+Y \leq 1) = f(0,0) + f(0,1) + f(1,0) + \int_0^1 \int_0^{1-x} \frac{3}{8} k \, dy \, dx = \frac{11}{16}.$$

4.2 Distribuciones condicionales

Dado un vector aleatorio bidimensional (X, Y) interesa conocer cómo se modifica la distribución de X si se conoce el valor que toma la variable Y (y recíprocamente). Entonces llamaremos distribución condicional de X a Y , notada X/Y a una nueva variable aleatoria que va a medir el comportamiento de X cuando conocemos Y . Hay que enfatizar que, en realidad, tendremos una distribución condicional $X/Y = y_0$ para cada valor observable y_0 de Y .

Caso discreto

Sea (X, Y) un vector aleatorio discreto con función de probabilidad conjunta:

$P_{ij} = P(x = i, y = j), \forall i \in I, j \in J$. Entonces si $Y = j_0, j_0 \in J$, la variable condicional $X/Y = j_0$ tiene distribución discreta con función de probabilidad:

$$P(X = i/Y = j_0) = \frac{P(X=i, Y=j_0)}{P(Y=j_0)}, \forall i \in I.$$

Simétricamente si $X = i_0$, entonces la variable condicional $Y/X = i_0$ tiene distribución discreta dada por:

$$P(Y = j/X = i_0) = \frac{P(X=i_0, Y=j)}{P(X=i_0)}, \forall j \in J.$$

Ejemplo 4.11:

Se lanzan dos dados. Sea X el resultado del primer dado e Y el máximo de los puntos obtenidos en ambos dados. Obtener la función de probabilidad conjunta, las marginales y las condicionales del vector aleatorio (X, Y) .

Probabilidad conjunta y marginales:

| XY | 1 | 2 | 3 | 4 | 5 | 6 | X |
|----|------|------|------|------|------|-------|------|
| 1 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 1/36 | 6/36 |
| 2 | 0 | 2/36 | 1/36 | 1/36 | 1/36 | 1/36 | 6/36 |
| 3 | 0 | 0 | 3/36 | 1/36 | 1/36 | 1/36 | 6/36 |
| 4 | 0 | 0 | 0 | 4/36 | 1/36 | 1/36 | 6/36 |
| 5 | 0 | 0 | 0 | 0 | 5/36 | 1/36 | 6/36 |
| 6 | 0 | 0 | 0 | 0 | 0 | 6/36 | 6/36 |
| Y | 1/36 | 3/36 | 5/36 | 7/36 | 9/36 | 11/36 | 1 |

Distribuciones de probabilidad condicionales:

| | |
|-------|-------------------------|
| X/Y=1 | 1 |
| | $\frac{1/36}{1/36} = 1$ |

| | | |
|-------|-----|-----|
| X/Y=2 | 1 | 2 |
| | 1/3 | 2/3 |

| | | | |
|-------|-----|-----|-----|
| X/Y=3 | 1 | 2 | 3 |
| | 1/5 | 1/5 | 3/5 |

| | | | | |
|-------|-----|-----|-----|-----|
| X/Y=4 | 1 | 2 | 3 | 4 |
| | 1/7 | 1/7 | 1/7 | 4/7 |

| | | | | | |
|-------|-----|-----|-----|-----|-----|
| X/Y=5 | 1 | 2 | 3 | 4 | 5 |
| | 1/9 | 1/9 | 1/9 | 1/9 | 5/9 |

| | | | | | | |
|-------|------|------|------|------|------|------|
| X/Y=6 | 1 | 2 | 3 | 4 | 5 | 6 |
| | 1/11 | 1/11 | 1/11 | 1/11 | 1/11 | 6/11 |

| | | | | | | |
|-------|-----|-----|-----|-----|-----|-----|
| Y/X=1 | 1 | 2 | 3 | 4 | 5 | 6 |
| | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

| | | | | | |
|-------|-----|-----|-----|-----|-----|
| Y/X=2 | 2 | 3 | 4 | 5 | 6 |
| | 2/6 | 1/6 | 1/6 | 1/6 | 1/6 |

| | | | | |
|-------|-----|-----|-----|-----|
| Y/X=3 | 3 | 4 | 5 | 6 |
| | 3/6 | 1/6 | 1/6 | 1/6 |

| | | | |
|-------|-----|-----|-----|
| Y/X=4 | 4 | 5 | 6 |
| | 4/6 | 1/6 | 1/6 |

| | | |
|-------|-----|-----|
| Y/X=5 | 5 | 6 |
| | 5/6 | 1/6 |

| | |
|-------|---|
| Y/X=6 | 6 |
| | 1 |

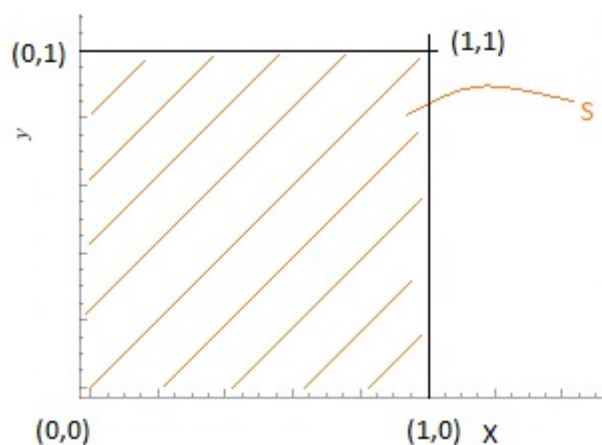
Caso absolutamente continuo

Sea (X, Y) un vector aleatorio absolutamente continuo con densidad conjunta $f(x, y)$ y marginales $f(x)$ y $f(y)$. Para cada valor y_0 con $f(y_0) > 0$, existe una variable condicional $X/Y = y_0$ que es absolutamente continua con densidad $f_{X/Y=y_0}(x) = \frac{f(x, y_0)}{f(y_0)}$. El recinto en el que varía x dependerá con frecuencia del valor de $Y = y_0$.

Simétricamente, Para cada valor x_0 con $f(x_0) > 0$, existe una variable condicional $Y/X = x_0$ que es absolutamente continua con densidad $f_{Y/X=x_0}(y) = \frac{f(x_0, y)}{f(x_0)}$. El recinto en el que varía Y dependerá con frecuencia del valor de $X = x_0$.

Ejemplo 4.12:

Sea (X, Y) un vector aleatorio bidimensional absolutamente continuo, que representa la selección aleatoria de un punto en el cuadrado unitario. Obtener las densidades condicionales.



$$f(x, y) = \begin{cases} 1 & \text{si } 0 \leq x, y \leq 1 \\ 0 & \text{fuera.} \end{cases}$$

$$f(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \int_0^1 1 dy = \begin{cases} 1 & \text{si } x \in [0, 1] \\ 0 & \text{fuera,} \end{cases} \rightsquigarrow U(0, 1).$$

$$f(y) = \int_{-\infty}^{+\infty} f(x, y) dx = \int_0^1 1 dx = \begin{cases} 1 & \text{si } y \in [0, 1] \\ 0 & \text{fuera,} \end{cases} \rightsquigarrow U(0, 1).$$

Si $y_o \in [0, 1]$,

$$f_{X/Y=y_o}(x) = \frac{f(x, y_o)}{f(y_o)} = \begin{cases} 1 & \text{si } x \in [0, 1] \\ 0 & \text{fuera,} \end{cases} \rightsquigarrow U(0, 1).$$

Si $x_o \in [0, 1]$,

$$f_{Y/X=x_o}(y) = \frac{f(x_o, y)}{f(x_o)} = \begin{cases} 1 & \text{si } y \in [0, 1] \\ 0 & \text{fuera,} \end{cases} \rightsquigarrow U(0, 1).$$

Ejemplo 4.13:

Sea (X, Y) un vector aleatorio bidimensional con densidad conjunta:

$$f(x, y) = \begin{cases} cx^2y & \text{si } x^2 \leq y \leq 1 \\ 0 & \text{fuera.} \end{cases}$$

Calcular las densidades condicionales:

$$\iint_{\mathbb{R}^2} f(x, y) \, dx dy = 1 = \int_{-1}^1 \int_{x^2}^1 cx^2y \, dy dx = 1; c = 21/4$$

$$f(x) = \int_{-\infty}^{+\infty} f(x, y) \, dy = \int_{x^2}^1 \frac{21}{4} x^2 y \, dy = \begin{cases} \frac{21}{8} x^2 (1 - x^4) & \text{si } x \in [-1, 1] \\ 0 & \text{fuera.} \end{cases}$$

$$f(y) = \int_{-\infty}^{+\infty} f(x, y) \, dx = \int_{-\sqrt{y}}^{+\sqrt{y}} \frac{21}{4} x^2 y \, dx = \begin{cases} 7y^2 \sqrt{y} & \text{si } y \in [0, 1] \\ 0 & \text{fuera.} \end{cases}$$

Si $y_0 \in [0, 1]$,

$$f_{X/Y=y_0}(x) = \frac{f(x, y_0)}{f(y_0)} = \begin{cases} \frac{3}{2} x^2 y_0^{-3/2} & \text{si } x \in [-\sqrt{y_0}, +\sqrt{y_0}] \\ 0 & \text{fuera.} \end{cases}$$

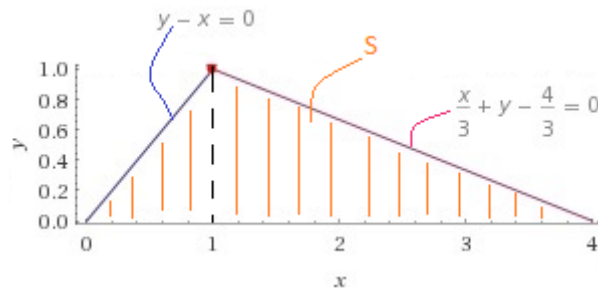
Si $x_0 \in [-1, 1]$,

$$f_{Y/X=x_0}(y) = \frac{f(x_0, y)}{f(x_0)} = \begin{cases} \frac{2y}{1-x_0^4} & \text{si } y \in [x_0^2, 1] \\ 0 & \text{fuera.} \end{cases}$$

Ejemplo 4.14:

Se elige un punto (x, y) al azar en el recinto S , triángulo con vértices en $(0, 0)$, $(1, 1)$, $(4, 0)$.
Obtener:

- La densidad conjunta.
- Marginales.
- Condicionales.



a) La densidad conjunta.

$$f(x,y) = \begin{cases} k & \text{si } (x,y) \in S \\ 0 & \text{fuera.} \end{cases}$$

$$\iint_S f(x,y) dx dy = 1 = \iint_S k dx dy = k \cdot \text{Area}(S) = 1; k = \frac{1}{2}$$

b) Marginales.

$$f(x) = \int_{-\infty}^{+\infty} f(x,y) dy = \begin{cases} \int_0^x \frac{1}{2} dy = \frac{x}{2} & \text{si } x \in [0, 1] \\ \int_0^{-\frac{x}{3} + \frac{4}{3}} \frac{1}{2} dy = -\frac{x}{6} + \frac{2}{3} & \text{si } x \in [1, 4] \\ 0 & \text{fuera.} \end{cases}$$

$$f(y) = \int_{-\infty}^{+\infty} f(x,y) dx = \int_y^{-3y+4} \frac{1}{2} dx = \begin{cases} -2y + 2 & \text{si } y \in [0, 1] \\ 0 & \text{fuera.} \end{cases}$$

c) Condicionales.

Si $y_0 \in (0, 1)$,

$$f_{X/Y=y_0}(x) = \begin{cases} \frac{\frac{1}{2}}{2-2y_0} = \frac{1}{4-4y_0} & \text{si } x \in [y_0, -3y_0+4] \\ 0 & \text{fuera.} \end{cases} \rightsquigarrow U(y_0, -3y_0+4).$$

Si $x_0 \in (0, 1)$,

$$f_{Y/X=x_0}(y) = \begin{cases} \frac{1}{x_0} & \text{si } y \in (0, x_0) \\ 0 & \text{fuera.} \end{cases} \rightsquigarrow U(0, x_0).$$

Si $x_0 \in (1, 4)$,

$$f_{Y/X=x_0}(y) = \begin{cases} \frac{1}{-\frac{x_0}{3} + \frac{4}{3}} = \frac{1}{-\frac{x_0}{3} + \frac{4}{3}} & \text{si } y \in (0, -\frac{x_0}{3} + \frac{4}{3}) \\ 0 & \text{fuera.} \end{cases} \rightsquigarrow U(0, -\frac{x_0}{3} + \frac{4}{3}).$$

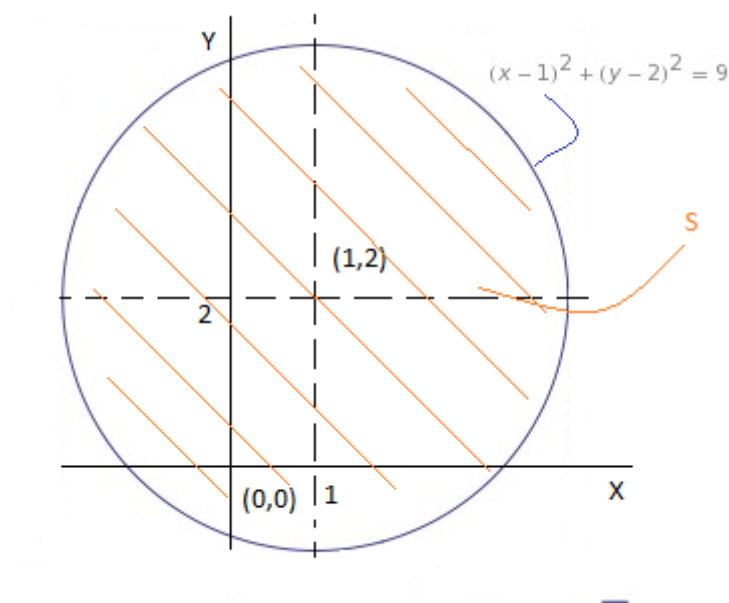
Ejemplo 4.15:

Se elige un punto al azar en el recinto S definido por :

$$S = \{(x,y) \in \mathbb{R}^2 / (x-1)^2 + (y-2)^2 \leq 9\}.$$

Calcular:

- La densidad conjunta.
- Marginales.
- Condicionales.



- La densidad conjunta.

$$f(x,y) = \begin{cases} k & \text{si } (x,y) \in S \\ 0 & \text{fuera.} \end{cases}$$

$$\iint_S f(x,y) dx dy = 1 = \iint_S k dx dy = k \cdot \text{Area}(S) = 1; k = \frac{1}{9\pi}$$

- Marginales.

$$f(x) = \int_{-\infty}^{+\infty} f(x,y) dy = \begin{cases} \int_{2-\sqrt{9-(x-1)^2}}^{2+\sqrt{9-(x-1)^2}} \frac{1}{9\pi} dy = \frac{2\sqrt{9-(x-1)^2}}{9\pi} & \text{si } x \in [-2, 4] \\ 0 & \text{fuera.} \end{cases}$$

$$f(y) = \int_{-\infty}^{+\infty} f(x,y) dx = \begin{cases} \int_{1-\sqrt{9-(y-2)^2}}^{1+\sqrt{9-(y-2)^2}} \frac{1}{9\pi} dy = \frac{2\sqrt{9-(y-2)^2}}{9\pi} & \text{si } y \in [-1, 5] \\ 0 & \text{fuera.} \end{cases}$$

- Condicionales.

Si $y_0 \in (-1, 5)$,

$$f_{X/Y=y_0}(x) = \begin{cases} \frac{1}{2\sqrt{9-(y_0-2)^2}} & \text{si } x \in [1 - \sqrt{9-(y_0-2)^2}, 1 + \sqrt{9-(y_0-2)^2}] \\ 0 & \text{fuera.} \end{cases}$$

Si $x_0 \in (-2, 4)$,

$$f_{Y/X=x_0}(y) = \begin{cases} \frac{1}{(2\sqrt{9-(x_0-1)^2})} & \text{si } y \in [2 - \sqrt{9-(x_0-1)^2}, 2 + \sqrt{9-(x_0-1)^2}] \\ 0 & \text{fuera.} \end{cases}$$

4.3 Independencia de variables aleatorias

Definición 4.6:

Diremos que dos variables aleatorias X e Y son independientes si y solo si $\forall A, B \in \sigma$ -álgebra de Borel en \mathbb{R} se verifica:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Podemos caracterizar la independencia en términos de la función de densidad en el caso absolutamente continuo y de la función de probabilidad en el caso discreto.

Proposición 4.1:

Sea (X, Y) un vector aleatorio bidimensional con función de distribución conjunta $F(x, y)$ y funciones de distribución marginales $F(x)$ y $F(y)$. Entonces X e Y son variables aleatorias independientes, si y solo si, $\forall x, y \in \mathbb{R}$, $F(x, y) = F(x)F(y)$.

Proposición 4.2:

Sea (X, Y) un vector aleatorio discreto con función de probabilidad conjunta $P(X = i, Y = j)$, $i \in I$, $j \in J$ y marginales $P(X = i), i \in I$, $P(Y = j), j \in J$. Entonces X e Y son independientes si y solo si $\forall i \in I, \forall j \in J$, $P(X = i, Y = j) = P(X = i) \cdot P(Y = j)$.

Desde el punto de vista intuitivo, la tabla de probabilidad conjunta debe ser el resultado de multiplicar los márgenes.

Proposición 4.3:

Sea (X, Y) un vector aleatorio absolutamente continuo con densidad conjunta $f(x, y)$ y marginales $f(x)$ y $f(y)$. Entonces X e Y son independientes si y solo si $\forall x, y \in \mathbb{R}$, $f(x, y) = f(x)f(y)$.

Proposición 4.4:

Si X e Y son variables independientes entonces:

$$X/Y=y_0 \equiv X \text{ y } Y/X=x_0 \equiv Y.$$

Demostración caso discreto:

$$P(X = x/Y=y_0) = \frac{P(X=x, Y=y_0)}{P(Y=y_0)} = \frac{P(X=x)P(Y=y_0)}{P(Y=y_0)} = P(X = x).$$

Demostración caso continuo:

$$f(X/Y=y_0) = \frac{f(x,y_0)}{f(y_0)} = \frac{f(x)f(y_0)}{f(y_0)} = f(x).$$

El resultado es lógico pues, bajo independencia, la información proporcionada por una variable no modifica las verosimilitudes de los valores de la otra.

Ejemplo 4.16:

Se lanzan dos dados. Sea X la suma de puntos obtenidos e Y la diferencia entre los puntos del primer dado y el segundo (con signo). Determinar si X e Y son variables aleatorias independientes.

| XY | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | X |
|----|------|------|------|------|------|------|------|------|------|------|------|------|
| 2 | 0 | 0 | 0 | 0 | 0 | 1/36 | 0 | 0 | 0 | 0 | 0 | 1/36 |
| 3 | 0 | 0 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 0 | 0 | 2/36 |
| 4 | 0 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 0 | 3/36 |
| 5 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 4/36 |
| 6 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 5/36 |
| 7 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 6/36 |
| 8 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 5/36 |
| 9 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 4/36 |
| 10 | 0 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 0 | 3/36 |
| 11 | 0 | 0 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 0 | 0 | 2/36 |
| 12 | 0 | 0 | 0 | 0 | 0 | 1/36 | 0 | 0 | 0 | 0 | 0 | 1/36 |
| Y | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 6/36 | 5/36 | 4/36 | 3/36 | 2/36 | 1/36 | 1 |

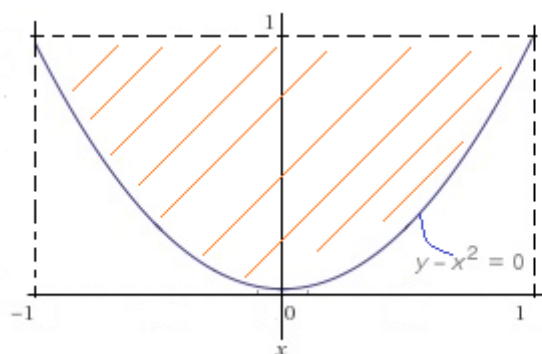
No son independientes dado que $P(X = x, Y = y) \neq P(X = x)P(Y = y)$. Ejemplo: $P(X = 2, Y = -5) = 0 \neq P(X = 2)P(Y = -5) = \frac{1}{36} \frac{1}{36}$.

Ejemplo 4.17:

Sea (X, Y) un vector aleatorio bidimensional con densidad conjunta.

$$f(x, y) = \begin{cases} kx^2(1-y) & \text{si } x^2 \leq y \leq 1 \\ 0 & \text{fuera.} \end{cases}$$

Determinar si X e Y son variables aleatorias independientes.



X e Y no pueden ser independientes ya que interdependen a través del recinto. Por lo tanto, si el recinto no es rectangular, las variables no son independientes. Nótese que, sin embargo, el

recíproco no es cierto. Es decir, aunque el recinto sea rectangular puede ocurrir que las variables no sean independientes. Por ejemplo el vector (X, Y) con

$$f(x, y) = \begin{cases} k(x+y) & \text{si } x, y \in [0, 1] \\ 0 & \text{fuera,} \end{cases}$$

tiene como soporte (conjunto de valores con densidad estrictamente positiva) un cuadrado pero X e Y no son independientes porque el polinomio $x+y$ no puede factorizarse.

4.4 Generalización al caso n-dimensional

Sea β^n la σ -álgebra de Borel en \mathbb{R}^n , es decir, la familia de conjuntos generados a partir de máximos, mínimos, uniones e intersecciones de los conjuntos de la forma $[-\infty, x_1]x[-\infty, x_2]x \dots x[-\infty, x_n]$ cuando $x_1, x_2, \dots, x_n \in \mathbb{R}$.

Definición 4.7:

Llamaremos vector aleatorio n-dimensional (X_1, X_2, \dots, X_n) a cualquier función definida en un espacio probabilístico $(\Omega, \sigma(\Omega), P)$ y con valores en \mathbb{R}^n :

$$\begin{aligned} (X_1, X_2, \dots, X_n) : (\Omega, \sigma(\Omega), P) &\rightarrow \mathbb{R}^n \\ \omega &\rightarrow (X_1(\omega), X_2(\omega), \dots, X_n(\omega)) \end{aligned}$$

y que verifique que para todo $x_1, x_2, \dots, x_n \in \mathbb{R}$,

$$\{\omega | X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_n(\omega) \leq x_n\} \equiv \{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\} \in \sigma(\Omega).$$

Proposición 4.5:

Si (X_1, X_2, \dots, X_n) es un vector aleatorio n-dimensional, entonces X_1, X_2, \dots, X_n son n variables aleatorias.

Como consecuencia es equivalente hablar de un vector aleatorio n-dimensional o de n variables aleatorias.

Función de distribución de un vector aleatorio n.-dimensional

Definición 4.8:

Dado un vector aleatorio n-dimensional (X_1, X_2, \dots, X_n) llamaremos función de distribución de tal vector a $F : \mathbb{R}^n \rightarrow \mathbb{R}$ que asigna a cada $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, $F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$.

Algunas propiedades:

$$i) F(x_1, x_2, \dots, x_n) \in [0, 1] \quad \forall x_1, x_2, \dots, x_n \in \mathbb{R}.$$

ii) $F(x_1, x_2, \dots, x_n)$ es no decreciente en cada variable, es decir, $\forall i = 1, 2, \dots, n$ si $x_i \leq x'_i$ entonces:

$$F(x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) \leq F(x_1, x_2, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n).$$

$$iii) \lim_{x_1 \rightarrow \infty} \lim_{x_2 \rightarrow \infty} \dots \lim_{x_n \rightarrow \infty} F(x_1, x_2, \dots, x_n) = 1.$$

$$iv) \lim_{x_1 \rightarrow -\infty} \lim_{x_2 \rightarrow -\infty} \dots \lim_{x_n \rightarrow -\infty} F(x_1, x_2, \dots, x_n) = 0.$$

Definición 4.9:

Diremos que el vector aleatorio n-dimensional (X_1, X_2, \dots, X_n) es discreto si X_1, X_2, \dots, X_n son discretas. En tal caso existe la llamada función de probabilidad conjunta del vector:

$$P(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n), i_1 \in I_1, i_2 \in I_2, \dots, i_n \in I_n$$

Esta función satisface las dos condiciones siguientes:

$$i) \sum_{i_1, i_2, \dots, i_n} P(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = 1$$

$$ii) P(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) \geq 0, \forall i_1, i_2, \dots, i_n.$$

Ejemplo 4.18:

Lanzamos tres monedas. Sea X_1, X_2, X_3 los resultados de los 3 lanzamientos. Entonces, llamando 0 a cara y 1 a cruz la función de probabilidad conjunta del vector (X_1, X_2, X_3) es:

$$P(X_1 = 0, X_2 = 0, X_3 = 0) = 1/8$$

$$P(X_1 = 0, X_2 = 0, X_3 = 1) = 1/8$$

.

.

.

$$P(X_1 = 1, X_2 = 1, X_3 = 1) = 1/8$$

Ejemplo 4.19:

Se extraen 4 cartas de una baraja española con remplazamiento. Sea X_1 el número de oros, X_2 el número de copas, X_3 el número de bastos y X_4 el número de espadas. La función de probabilidad conjunta es:

$$P(X_1 = i, X_2 = j, X_3 = k, X_4 = l) = \begin{cases} \frac{4!}{i!j!k!l!} \left(\frac{1}{4}\right)^i \left(\frac{1}{4}\right)^j \left(\frac{1}{4}\right)^k \left(\frac{1}{4}\right)^l & \text{si } i, j, k, l \geq 0, i + j + k + l = 4 \\ 0 & \text{fuera} \end{cases}$$

$$= \begin{cases} \frac{4!}{i!j!k!l!} \left(\frac{1}{4}\right)^4 & \text{si } i, j, k, l \geq 0, i + j + k + l = 4 \\ 0 & \text{fuera} \end{cases}$$

Ejemplo 4.20:

Sea (X_1, X_2, X_3) el vector tridimensional, número de unos, número de doses y número de treses al lanzar 3 dados. Su función de probabilidad conjunta es:

$$P(X_1 = i, X_2 = j, X_3 = k) = \begin{cases} \frac{3!}{i!j!k!(3-i-j-k)!} \left(\frac{1}{6}\right)^{i+j+k} \left(\frac{1}{2}\right)^{3-i-j-k} & \text{si } i, j, k \geq 0, i+j+k \leq 3 \\ 0 & \text{fuera} \end{cases}$$

Definición 4.10:

Diremos que (X_1, X_2, \dots, X_n) es un vector aleatorio absolutamente continuo cuando su función de distribución $F(x_1, x_2, \dots, x_n)$ es continua, es decir:

$$\lim_{x_1 \rightarrow a_1} \lim_{x_2 \rightarrow a_2} \dots \lim_{x_n \rightarrow a_n} F(x_1, x_2, \dots, x_n) = F(a_1, a_2, \dots, a_n) \in \mathbb{R}.$$

Entonces existe una función $f(x_1, x_2, \dots, x_n)$ llamada función de densidad tal que:

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f(t_1, t_2, \dots, t_n) dt_1 dt_2 \dots dt_n.$$

En consecuencia:

$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F(x_1, x_2, \dots, x_n).$$

Entonces, $\forall A \subseteq \mathbb{R}^n$, $P(A) = \iint \dots \int_A f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$.

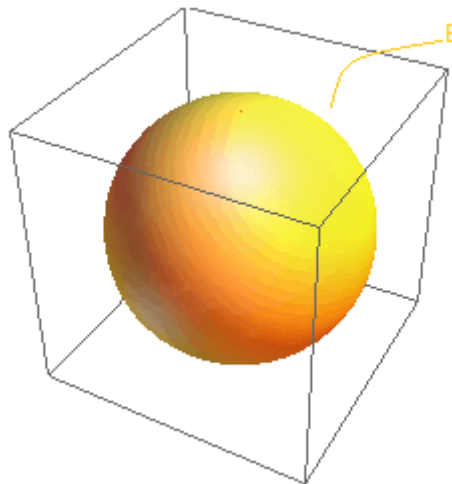
$f(x_1, x_2, \dots, x_n)$ Satisface:

$$i) f(x_1, x_2, \dots, x_n) \geq 0, \forall x_1, \dots, x_n \in \mathbb{R}$$

$$ii) \iint \dots \int_{\mathbb{R}^n} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = 1.$$

Ejemplo 4.21:

Sea (X, Y, Z) el vector aleatorio tridimensional absolutamente continuo con densidad uniforme en la esfera de radio 1.



$$f(x, y, z) = \begin{cases} k & \text{si } (x, y, z) \in E \\ 0 & \text{fuera} \end{cases}$$

Dado que $1 = \iiint_E k \, dx dy dz = k \cdot \text{volumen}(E) = k \cdot \frac{4\pi}{3} \Rightarrow k = \frac{3}{4\pi}$

Distribuciones Marginales

En el caso n-dimensional el concepto de distribución marginal es mas amplio que en el caso bivalente ya que aquí se incluirán las distribuciones de todos los subvectores y no solo de los univariantes. Es decir, llamaremos distribución marginal de (X_1, X_2, \dots, X_n) a cualquier subvector $(X_{i_1}, X_{i_2}, \dots, X_{i_r})$ con $i_1, i_2, \dots, i_r \in \{1, 2, \dots, n\}$ y $r < n$.

Caso discreto. Si el vector (X_1, X_2, \dots, X_n) es un vector discreto y $P(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n)$ para $k_1 \in I_1, k_2 \in I_2, \dots, k_n \in I_n$ es la función de probabilidad conjunta, para obtener la función de probabilidad del vector marginal $(X_{i_1}, X_{i_2}, \dots, X_{i_r})$ sumaremos en todos los valores de todas las variables diferentes de $X_{i_1}, X_{i_2}, \dots, X_{i_n}$

Ejemplo: 4.22

Sea (X_1, X_2, X_3) el vector “número de veces que salen el 1, el 2 y el 3 al lanzar 3 dados”. La función de probabilidad conjunta es:

$$P(X_1 = i, Y = j, Z = k) = \begin{cases} \frac{3!}{i!j!k!(3-i-j-k)!} \left(\frac{1}{6}\right)^{i+j+k} \left(\frac{1}{2}\right)^{3-i-j-k} & \text{si } i, j, k \geq 0, i+j+k \leq 3 \\ 0 & \text{fuera} \end{cases}$$

Marginales

$$\begin{aligned} P(X = i) &= \sum_{j,k} P(X_1 = i, X_2 = j, X_3 = k) = \sum_{j,k} \frac{3!}{i!j!k!(3-i-j-k)!} \left(\frac{1}{6}\right)^{i+j+k} \left(\frac{1}{2}\right)^{3-i-j-k} \\ &= \left(\frac{1}{6}\right)^i \left(\frac{1}{2}\right)^{3-i} \frac{1}{i!} \sum_{j,k} \frac{3!}{j!k!(3-i-j-k)!} \left(\frac{1}{3}\right)^j \left(\frac{1}{3}\right)^k \\ &= \left(\frac{1}{6}\right)^i \left(\frac{1}{2}\right)^{3-i} \frac{1}{i!} \left(\frac{3!}{(3-i)!}\right) \sum_{j,k} \frac{(3-i)!}{j!k!(3-i-j-k)!} \left(\frac{1}{3}\right)^j \left(\frac{1}{3}\right)^k \\ &= \left(\frac{1}{6}\right)^i \left(\frac{1}{2}\right)^{3-i} \frac{1}{i!} \left(\frac{3!}{(3-i)!}\right) \left(\frac{1}{3} + \frac{1}{3} + 1\right)^{3-i} = \binom{3}{i} \left(\frac{5}{6}\right)^{3-i} \left(\frac{1}{6}\right)^i \sim \text{Bin}\left(3, \frac{1}{6}\right) \end{aligned}$$

Y, por simetría:

$$\begin{aligned} Y &\sim \text{Bin}\left(3, \frac{1}{6}\right), \\ Z &\sim \text{Bin}\left(3, \frac{1}{6}\right). \end{aligned}$$

Caso continuo. Si (X_1, \dots, X_n) es un vector absolutamente continuo con densidad $f(x_1, \dots, x_n)$ entonces la densidad de cualquier subvector marginal $(X_{i_1}, \dots, X_{i_r})$ se obtendrá integrando en las restantes variables.

Independencia

Dado un vector aleatorio n -dimensional (X_1, X_2, \dots, X_n) , diremos que las variables X_1, X_2, \dots, X_n son variables independientes si y solo si $\forall B_1, B_2, \dots, B_n \subseteq \beta$ (σ -álgebra de Borel en \mathbb{R}) se verifica:

$$P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = \prod_{i=1}^n P(X_i \in B_i)$$

Caracterización de la independencia

Proposición 4.6:

Sea (X_1, X_2, \dots, X_n) un vector aleatorio n -dimensional con función de distribución conjunta $F(x_1, x_2, \dots, x_n)$ y marginales $F(x_1), F(x_2), \dots, F(x_n)$. Entonces las variables X_1, X_2, \dots, X_n son independientes si y solo si

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F(x_i), \forall x_1, x_2, \dots, x_n \in \mathbb{R}.$$

Proposición 4.7:

Sea (X_1, X_2, \dots, X_n) un vector aleatorio n -dimensional discreto con función de probabilidad conjunta $P(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n), i_1 \in I_1, \dots, i_n \in I_n$ y marginales $P(X_1 = i_1), i_1 \in I_1, \dots, P(X_n = i_n), i_n \in I_n$. Entonces X_1, X_2, \dots, X_n son variables aleatorias independientes si y solo si

$$P(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = \prod_{j=1}^n P(X_j = i_j), i_1 \in I_1, \dots, i_n \in I_n$$

Proposición 4.8:

Sean X_1, X_2, \dots, X_n variables absolutamente continuas con densidad conjunta $f(x_1, x_2, \dots, x_n)$. Entonces X_1, X_2, \dots, X_n son variables aleatorias independientes si y solo si

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i) \forall x_1, x_2, \dots, x_n \in \mathbb{R}$$

Ejemplo 4.22:

Se lanzan de tres monedas. Sea $X_i = \begin{cases} 0 & \text{si Cruz} \\ 1 & \text{si Cara,} \end{cases} i = 1, 2, 3.$

$$P(X_1 = a, X_2 = b, X_3 = c) = 1/8, \forall a, b, c \in \{0, 1\}$$

luego,

$$P(X_1 = a) = \sum_{b,c} P(X_1 = a, X_2 = b, X_3 = c) = 4 \frac{1}{8} = \frac{1}{2}.$$

Entonces,

$$X_1 = \begin{cases} 0 & 1/2 \\ 1 & 1/2 \end{cases}$$

Y por simetría:

$$X_2 = \begin{cases} 0 & 1/2 \\ 1 & 1/2 \end{cases}$$

$$X_3 = \begin{cases} 0 & 1/2 \\ 1 & 1/2 \end{cases}$$

X_1 , X_2 y X_3 son independientes.

Ejemplo 4.23:

Sea (X_1, X_2, X_3) un vector absolutamente continuo con densidad:

$$f(x_1, x_2, x_3) = \begin{cases} e^{-x_1-x_2-x_3} & \text{si } x_1, x_2, x_3 \geq 0 \\ 0 & \text{fuera.} \end{cases}$$

Entonces,

$$f(x_1) = \begin{cases} \int_0^\infty \int_0^\infty e^{-x_1-x_2-x_3} dx_2 dx_3 = e^{-x_1} \int_0^\infty e^{-x_2} dx_2 \int_0^\infty e^{-x_3} dx_3 = e^{-x_1} & \text{si } x_1 \geq 0 \\ 0 & \text{fuera} \end{cases}$$

Por simetría:

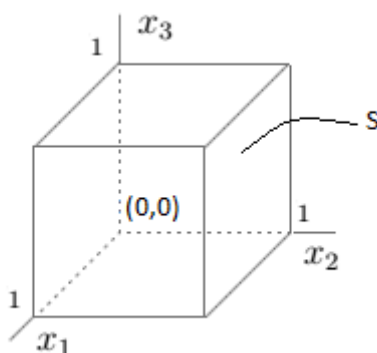
$$f(x_2) = \begin{cases} e^{-x_2} & \text{si } x_2 \geq 0 \\ 0 & \text{fuera} \end{cases}$$

$$f(x_3) = \begin{cases} e^{-x_3} & \text{si } x_3 \geq 0 \\ 0 & \text{fuera} \end{cases}$$

Por tanto, $f(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3) \forall x_1, x_2, x_3$, por lo que podemos afirmar que X_1, X_2 y X_3 son variables aleatorias independientes.

Ejemplo 4.24:

Sea (X_1, X_2, \dots, X_n) un punto seleccionado al azar en el cubo $0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1$.



Entonces,

$$f(x_1, x_2, x_3) = \begin{cases} 1 & \text{si } (x_1, x_2, x_3) \in S \\ 0 & \text{fuera.} \end{cases}$$

$$f(x_1) = \begin{cases} \int_0^1 \int_0^1 1 \, dx_2 dx_3 = 1 & \text{si } x_1 \in (0, 1) \\ 0 & \text{fuera} \end{cases}$$

Por simetría

$$f(x_2) = \begin{cases} 1 & \text{si } x_2 \in (0, 1) \\ 0 & \text{fuera} \end{cases}$$

$$f(x_3) = \begin{cases} 1 & \text{si } x_3 \in (0, 1) \\ 0 & \text{fuera} \end{cases}$$

Por tanto, $f(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3) \forall x_1, x_2, x_3$, por lo que X_1, X_2 y X_3 son variables aleatorias independientes.

Bibliografía

1. A.A. Borovkov (2013). *Probability Theory*. Springer.
2. A. Klenke (2014). *Probability Theory. A Comprehensive Course*. Springer.
3. M.H. DeGroot (1988). *Probability and Statistics*. Wilmington, Delaware, E.U.A. Addison-Wesley Iberoamericana, S.A.
4. V.K. Rohatgi and A.K. Md. Ehsanes Saleh (2001). *An Introduction to Probability and Statistics*. John Wiley and Sons.

Tema 5

Esperanza de vectores aleatorios. Esperanza de sumas y productos de variables aleatorias. Covarianza. Correlación. Transformaciones lineales de variables aleatorias.

Definición 5.1:

Dado un v.a. n-dimensional (X_1, X_2, \dots, X_n) llamaremos vector de medias o vector de esperanzas al vector $(EX_1, EX_2, \dots, EX_n)$, donde

$$EX_i = \begin{cases} \sum_{j \in J} jP(X_i = j) & \text{si } X_i \text{ es una v.a. discreta} \\ \int_{-\infty}^{\infty} x_i f(x_i) dx & \text{si } X_i \text{ es una v.a. continua.} \end{cases}$$

Ejemplo 5.1:

Se lanzan dos dados. Sea X el mínimo de los puntos obtenidos en ambos dados e Y el máximo.

| X/Y | 1 | 2 | 3 | 4 | 5 | 6 | X |
|-----|------|------|------|------|------|-------|-------|
| 1 | 1/36 | 2/36 | 2/36 | 2/36 | 2/36 | 2/36 | 11/36 |
| 2 | 0 | 1/36 | 2/36 | 2/36 | 2/36 | 2/36 | 9/36 |
| 3 | 0 | 0 | 1/36 | 2/36 | 2/36 | 2/36 | 7/36 |
| 4 | 0 | 0 | 0 | 1/36 | 2/36 | 2/36 | 5/36 |
| 5 | 0 | 0 | 0 | 0 | 1/36 | 2/36 | 3/36 |
| 6 | 0 | 0 | 0 | 0 | 0 | 1/36 | 1/36 |
| Y | 1/36 | 3/36 | 5/36 | 7/36 | 9/36 | 11/36 | 1 |

$$EX = 1 \cdot \frac{11}{36} + 2 \cdot \frac{9}{36} + \dots + 6 \cdot \frac{1}{36} = \frac{9}{4}$$

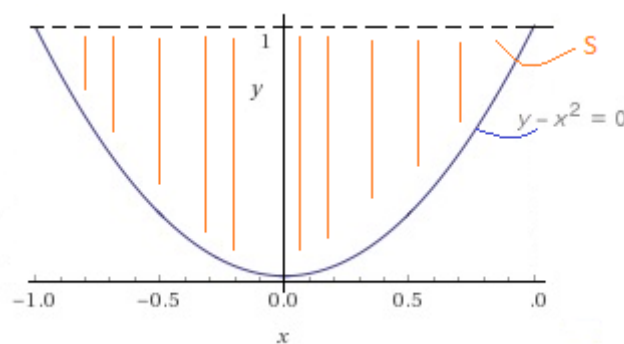
$$EY = 1 \cdot \frac{1}{36} + 2 \cdot \frac{3}{36} + \dots + 6 \cdot \frac{11}{36} = \frac{161}{36}$$

$$(EX, EY) = \left(\frac{9}{4}, \frac{161}{36} \right)$$

Ejemplo 5.2:

Sea (X, Y) un vector aleatorio bidimensional con densidad conjunta.

$$f(x, y) = \begin{cases} cx^2y & \text{si } x^2 \leq y \leq 1 \\ 0 & \text{fuera} \end{cases}$$



$$\iint_{\mathbb{R}^2} f(x, y) \, dx dy = 1 = \int_{-1}^1 \int_{x^2}^1 c x^2 y \, dy dx = 1; c = 21/4$$

$$f(x) = \begin{cases} \frac{21}{8} x^2 (1 - x^4) & \text{si } x \in [-1, 1] \\ 0 & \text{fuera.} \end{cases}$$

$$f(y) = \begin{cases} \frac{7}{2} y^2 \sqrt{y} & \text{si } y \in [0, 1] \\ 0 & \text{fuera.} \end{cases}$$

$$EX = \int x \cdot f(x) dx = \int_{-1}^1 x \left(\frac{21}{8} x^2 (1 - x^4) \right) dx = 0$$

$$EY = \int y \cdot f(y) dy = \int_0^1 y (7 y^2 \sqrt{y}) dy = \frac{7}{9}$$

$$(EX, EY) = (0, \frac{7}{9})$$

5.1 Propiedades de la Esperanza Matemática

i) Si $g(X, Y)$ es función de las variables aleatorias X e Y entonces:

$$E(g(X, Y)) = \begin{cases} \iint g(x, y) f(x, y) \, dx dy & \text{si } X, Y \text{ continuas} \\ \sum_i \sum_j g(i, j) P(X = i, Y = j) & \text{si } X, Y \text{ discretas.} \end{cases}$$

ii) La esperanza es lineal, es decir, $\forall a, b \in \mathbb{R}$ y $\forall X, Y$ variables aleatorias, $E(aX + bY) = aEX + bEY$.

Demostración:

$$E(aX + bY) = \iint (ax + by) f(x, y) dx dy = a \iint x f(x, y) dy dx + b \iint y f(x, y) dy dx = a \int x f(x) dx + b \int y f(y) dy = aEX + bEY.$$

El resultado se generalice sin dificultad (por inducción) para X_1, \dots, X_n variables aleatorias con esperanza finita y $a_1, \dots, a_n \in \mathbb{R}$, obteniendo que

$$E(a_1X_1 + \dots + a_nX_n) = a_1EX_1 + \dots + a_nEX_n.$$

iii) Si X e Y son variables aleatorias independientes entonces $EXY = EXEY$.

Demostración:

$$EXY = \iint xyf(x,y) dx dy = \iint xyf(x)f(y) dx dy = \int xf(x) dx \int yf(y) dy = EXEY.$$

Ejemplo 5.3:

Retomando el ejemplo anterior, donde (X, Y) es un vector aleatorio bidimensional con densidad conjunta:

$$f(x,y) = \begin{cases} \frac{21}{4}x^2y & \text{si } x^2 \leq y \leq 1 \\ 0 & \text{fuera,} \end{cases}$$

y el vector de esperanzas es $(EX, EY) = (0, \frac{7}{9})$, entonces,

$$E(2X + 3Y) = 2EX + 3EY = 2(0) + 3\frac{7}{9} = \frac{7}{3}$$

5.2 Momentos de un vector bidimensional respecto a un punto, respecto al origen y respecto a la media

Definición 5.2:

Dado un vector aleatorio bidimensional (X, Y) llamaremos momentos de orden r, s respecto al vector $(a, b) \in \mathbb{R}^2$ a:

$$\alpha_{r,s}(a, b) = E[(x - a)^r \cdot (Y - b)^s].$$

En particular si $a = 0, b = 0$, $\alpha_{r,s}(0, 0)$ recibe el nombre de momento de orden (r, s) respecto del origen.

$$\alpha_{r,s}(0, 0) = E(X^r \cdot Y^s)$$

Y si $a = EX, b = EY$, $\alpha_{r,s}(EX, EY)$ recibe el nombre de momento de orden (r, s) respecto de la media.

$$\alpha_{r,s}(EX, EY) = E((X - EX)^r \cdot (Y - EY)^s).$$

Los momentos respecto a la media se pueden expresar siempre en función de los momentos respecto al origen.

Ejemplo 5.4:

$$\begin{aligned} \alpha_{2,0}(EX, EY) &= E((X - EX)^2) = E(X^2 + E^2X - 2X \cdot EX) = EX^2 + E^2X - 2EXEX = EX^2 - E^2X \\ &= \alpha_{2,0}(0, 0) - \alpha_{1,0}^2(0, 0). \end{aligned}$$

Definición 5.3:

Dado un vector aleatorio (X, Y) se llama covarianza de (X, Y) , notada $Cov(X, Y)$, al momento de orden $(1, 1)$ respecto a la media:

$$Cov(X, Y) = \alpha_{1,1}(EX, EY) = E[(X - EX)(Y - EY)]$$

Proposición 5.1:

$Cov(X, Y) \equiv EXY - EXEY$ para todo vector (X, Y) siempre y cuando exista tanto EXY como EX y EY .

Demostración:

$$\begin{aligned} Cov(X, Y) &= E[(X - EX)(Y - EY)] = E[XY - XEY - YEX + EXEY] \\ &= EXY - EXEY = \alpha_{1,1}(0, 0) - \alpha_{1,0}(0, 0) \cdot \alpha_{0,1}(0, 0) \end{aligned}$$

Propiedades de la covarianza:

i) La covarianza es simétrica, es decir, $Cov(X, Y) = Cov(Y, X)$.

Demostración:

Demostración trivial a partir de la definición.

ii) Si X e Y son variables aleatorias independientes entonces su covarianza es nula.

Demostración: Es trivial pues si X e Y son independientes, $EXY = EXEY$. Entonces,

$$Cov(X, Y) = EXY - EXEY = EXEY - EXEY = 0$$

iii) El recíproco no es cierto, existen variables con covarianza nula y que no son independientes.

Contraejemplo Sean X e Y la suma y diferencia de puntos al lanzar dos dados.

| X/Y | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | X. |
|-----|------|------|------|------|------|------|------|------|------|------|------|------|
| 2 | 0 | 0 | 0 | 0 | 0 | 1/36 | 0 | 0 | 0 | 0 | 0 | 1/36 |
| 3 | 0 | 0 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 0 | 0 | 2/36 |
| 4 | 0 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 0 | 3/36 |
| 5 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 4/36 |
| 6 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 5/36 |
| 7 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 6/36 |
| 8 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 5/36 |
| 9 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 4/36 |
| 10 | 0 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 0 | 3/36 |
| 11 | 0 | 0 | 0 | 0 | 1/36 | 0 | 1/36 | 0 | 0 | 0 | 0 | 2/36 |
| 12 | 0 | 0 | 0 | 0 | 0 | 1/36 | 0 | 0 | 0 | 0 | 0 | 1/36 |
| Y. | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 6/36 | 5/36 | 4/36 | 3/36 | 2/36 | 1/36 | 1 |

$$EXY = \sum_x \sum_y xyP(X=x, Y=y) = 2 \cdot 0 \cdot \frac{1}{36} + 3 \cdot (-1) \cdot \frac{1}{36} + 3 \cdot 1 \cdot \frac{1}{36} + \dots + 12 \cdot 0 \cdot \frac{1}{36} = 0$$

$$EY = \sum_y yP(Y=y) = -5 \cdot \frac{1}{36} + (-4) \cdot \frac{2}{36} + \dots + 4 \cdot \frac{2}{36} + 5 \cdot \frac{1}{36} = 0$$

$Cov(X, Y) = EXY - EXEY = 0 - 0 = 0$ y sin embargo las variables no son independientes.

iv) La covarianza es bilineal, es decir, si $a, b, c, d, e, f \in \mathbb{R}$ y X_1, X_2, X_3, X_4 son variables aleatorias, entonces la covarianza $Cov(aX_1 + bX_2 + c, dX_3 + eX_4 + f) = adCov(X_1, X_3) + aeCov(X_1, X_4) + bdCov(X_2, X_3) + beCov(X_2, X_4)$.

Demostración: $Cov(aX_1 + bX_2 + c, dX_3 + eX_4 + f) = E(aX_1 + bX_2 + c)(dX_3 + eX_4 + f) - E(aX_1 + bX_2 + c)E(dX_3 + eX_4 + f) = E(adX_1X_3 + aeX_1X_4 + afX_1 + bdX_2X_3 + beX_2X_4 + bfX_2 + cdX_3 + ceX_4 + cf) - E(aX_1 + bX_2 + c)E(dX_3 + eX_4 + f) = ad(EX_1X_3 - EX_1EX_3) + ae(EX_1X_4 - EX_1EX_4) + bd(EX_2X_3 - EX_2EX_3) + be(EX_2X_4 - EX_2EX_4) = adCov(X_1, X_3) + aeCov(X_1, X_4) + bdCov(X_2, X_3) + beCov(X_2, X_4)$

Ejemplo 5.5:

Si $Cov(X_1, X_3) = -2$, $Cov(X_1, X_4) = 1$, $Cov(X_2, X_3) = 0$ y $Cov(X_2, X_4) = -2$, entonces,

$$\begin{aligned} Cov(3X_1 + 4X_2 + 1, -2X_3 - 3X_4 + 5) &= 3 \cdot (-2)Cov(X_1, X_3) + 3 \cdot (-3)Cov(X_1, X_4) \\ &\quad + 4 \cdot (-2)Cov(X_2, X_3) + 4 \cdot (-3)Cov(X_2, X_4) = 27 \end{aligned}$$

v) \forall variable X , $Cov(X, X) = Var(X)$.

Demostración: $Cov(X, X) = EXX - EXEX = EX^2 - E^2X = Var(X)$.

$$vi) \text{Var}(aX + bY) = \text{Cov}(aX + bY, aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y).$$

Demostración:

$$\text{Var}(aX + bY) = \text{Cov}(aX + bY, aX + bY) \text{ que usando la propiedad de bilateralidad anterior, } a^2\text{Cov}(X, Y) + b^2\text{Cov}(X, Y) + ab\text{Cov}(X, Y) + ba\text{Cov}(X, Y) = a^2\text{Var}(x) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y).$$

vii) La covarianza puede tomar valores positivos, negativos y nulos. La covarianza es positiva cuando al aumentar el valor de una variable también aumenta el valor de la otra y toma valores negativos en el caso contrario.

Ejemplo 5.6:

Se extraen 3 cartas de una baraja española con remplazamiento. Sea X_1 el número de oros y X_2 el número de copas. Calcular la $\text{Cov}(X_1, X_2)$.

$$P(X_1 = i, X_2 = j) = \begin{cases} \frac{3!}{i!j!(3-i-j)!} \left(\frac{1}{4}\right)^i \left(\frac{1}{4}\right)^j \left(\frac{1}{2}\right)^{3-i-j} & \text{si } i, j \geq 0, i+j \leq 3 \\ 0 & \text{fuera.} \end{cases}$$

$$E(X_1 X_2) = 0 \cdot 0 \cdot P(X_1 = 0, X_2 = 0) + 0 \cdot 1 \cdot P(X_1 = 0, X_2 = 1) + \dots + 3 \cdot 0 \cdot P(X_1 = 3, X_2 = 0) = 3/8$$

$$X_1 \sim B(3, 1/4), X_2 \sim B(3, 1/4) \Rightarrow EX_1 = EX_2 = 3/4$$

$$\text{Cov}(X_1, X_2) = EX_1 X_2 - EX_1 EX_2 = -\frac{3}{16} \leq 0.$$

La covarianza debe ser negativa, pues cuantos más oros se obtengan menos copas se podrán obtener.

Ejemplo 5.7:

Una secretaria escribe n cartas y las direcciones en n sobres, luego introduce las cartas en los sobres al azar. Calcular la probabilidad de que al menos una carta llegue a su destinatario.

Sea,

$$X_i = \begin{cases} 1 & \text{si la carta } i \text{ llega a su destinatario} \\ 0 & \text{si la carta } i \text{ no llega a su destinatario.} \end{cases}$$

Calcular la $\text{Cov}(X_i, X_j) \forall i \neq j$.

Sea A_i el suceso "la carta i -ésima llega a su destinatario", pudiendo llegar también otras cartas. Por tanto A_1, A_2, \dots, A_n no son sucesos mutuamente excluyentes. El suceso que al menos una carta llegue a su destinatario es $\cup_{i=1}^n A_i$. Para calcular su probabilidad usaremos el llamado principio de inclusión-exclusión.

Primero se suman las probabilidades de los n sucesos individuales. Luego, se resta la suma de las probabilidades de las intersecciones de todos los pares posibles del sucesos (que serán $\binom{n}{2}$), posteriormente se suman las probabilidades de las intersecciones de tres sucesos (de los

cuales hay $\binom{n}{3}$ y así sucesivamente hasta que finalmente se suma o se resta la probabilidad de la intersección de los n sucesos, dependiendo de si n es un número impar o par, es decir, el último término sera: $(-1)^{(n+1)}P(A_1, A_2, \dots, A_n)$. Entonces,

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n).$$

Se tiene que: para $i = 1, \dots, n$, $P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$. Para $i < j$, $i, j \in \{1, \dots, n\}$, $P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$. Para $i < j < k$, $i, j, k \in \{1, \dots, n\}$, $P(A_i \cap A_j \cap A_k) = \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)}$ y así sucesivamente. Entonces,

$$\begin{aligned} P(\cup_{i=1}^n A_i) &= n \frac{1}{n} - \binom{n}{2} \frac{1}{n(n-1)} + \binom{n}{3} \frac{1}{n(n-1)(n-2)} \\ &+ \dots + (-1)^{n+1} \frac{1}{n!} = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^n \frac{1}{n!} \xrightarrow{n \rightarrow \infty} 1 - e^{-1} = 0,6321 \end{aligned}$$

La serie converge muy rápidamente por lo que la probabilidad no depende (apenas) de n .

Para calcular $Cov(X_i, X_j)$ necesitamos conseguir la distribución de X_i , la de X_j y la de $X_i X_j$.

$$X_i = \begin{cases} 1 & P(A_i) = \frac{1}{n} \\ 0 & \text{en otro caso,} \end{cases}$$

análogo para X_j . Entonces, utilizando los resultados anteriores, la distribución de X_i , $i = 1, \dots, n$, y la distribución de

$$X_i \cdot X_j = \begin{cases} 1 & P(A_i \cap A_j) \\ 0 & 1 - P(A_i \cap A_j) \end{cases} = \begin{cases} 1 & \frac{1}{n(n-1)} \\ 0 & 1 - \frac{1}{n(n-1)}, \end{cases}$$

de donde $E(X_i \cdot X_j) = \frac{1}{n(n-1)}$ y $E(X_i) = E(X_j) = \frac{1}{n}$. Luego,

$$Cov(X_i, X_j) = E(X_i \cdot X_j) - E(X_i)E(X_j) = \frac{1}{n(n-1)} - \frac{1}{n} \frac{1}{n} = \frac{1}{n^2(n-1)} \geq 0.$$

5.3 Correlación lineal entre variables

En probabilidad y estadística, la correlación lineal indica la intensidad y la dirección de una relación lineal entre dos variables aleatorias. Se considera que dos variables cuantitativas están linealmente correlacionadas cuando los valores de una de ellas se pueden obtener de manera aproximada como una función lineal de los valores de la otra. La correlación entre dos variables no implica, por sí misma, ninguna relación de causalidad.

Definición 5.4:

Dado un vector aleatorio (X, Y) llamaremos coeficiente de correlación lineal de X e Y notado $\rho(X, Y)$ a:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

5.3.1 Propiedades del coeficiente de correlación

El coeficiente de correlación verifica las siguientes propiedades.

i) Sean $a, b, c, d \in \mathbb{R}$ y sean X e Y variables aleatorias, entonces $|\rho(aX + b, cY + d)| = \rho(X, Y)$. En particular, si $ac > 0$, $\rho(aX + b, cY + d) = \rho(X, Y)$, y si $ac < 0$, $\rho(aX + b, cY + d) = -\rho(X, Y)$.

Demostración:

$$\begin{aligned}\rho(aX + b, cY + d) &= \frac{\text{Cov}(aX + b, cY + d)}{\sqrt{\text{Var}(aX + b)}\sqrt{\text{Var}(cY + d)}} = \frac{ac\text{Cov}(X, Y)}{\sqrt{a^2\text{Var}(X)}\sqrt{c^2\text{Var}(Y)}} \\ &= \frac{ac\text{Cov}(X, Y)}{|a|\sqrt{\text{Var}(X)}|c|\sqrt{\text{Var}(Y)}} = \pm\rho(X, Y),\end{aligned}$$

y sera positivo si $ac > 0$; por el contrario, sera negativo si $ac < 0$.

ii) $\forall X, Y$ variables aleatorias, $-1 \leq \rho(X, Y) \leq 1$.

Demostración:

A partir de X e Y definimos dos nuevas variables, las tipificaciones de X e Y

$$X^* = \frac{X - EX}{\sqrt{\text{Var}(X)}}, Y^* = \frac{Y - EY}{\sqrt{\text{Var}(Y)}},$$

que verifica que $EX^* = EY^* = 0$ y $\text{Var}X^* = \text{Var}Y^* = 1$ Por la propiedad i) sabemos que $\rho(X, Y) = \rho(X^*, Y^*)$ ya que en este caso:

$a = \frac{1}{\sqrt{\text{Var}(X)}}$ y $c = \frac{1}{\sqrt{\text{Var}(Y)}}$ y por tanto $ac > 0$,

$$\rho(X^*, Y^*) = \frac{\text{Cov}(X^*, Y^*)}{\sqrt{\text{Var}(X^*)}\sqrt{\text{Var}(Y^*)}} = \frac{\text{Cov}(X^*, Y^*)}{1 \cdot 1} = \text{Cov}(X^*, Y^*)$$

Además

$$\text{Var}(X^* + Y^*) = \text{Var}(X^*) + \text{Var}(Y^*) + 2\text{Cov}(X^*, Y^*) = 2 + 2\text{Cov}(X^*, Y^*),$$

Entonces,

$$0 \leq 2 + 2\text{Cov}(X^*, Y^*) \Rightarrow \text{Cov}(X^*, Y^*) \geq -1$$

Paralelamente,

$$\text{Var}(X^* - Y^*) = \text{Var}(X^*) + \text{Var}(Y^*) - 2\text{Cov}(X^*, Y^*) = 2 - 2\text{Cov}(X^*, Y^*).$$

Por lo que,

$$0 \leq 2 - 2\text{Cov}(X^*, Y^*) \Rightarrow \text{Cov}(X^*, Y^*) \leq 1.$$

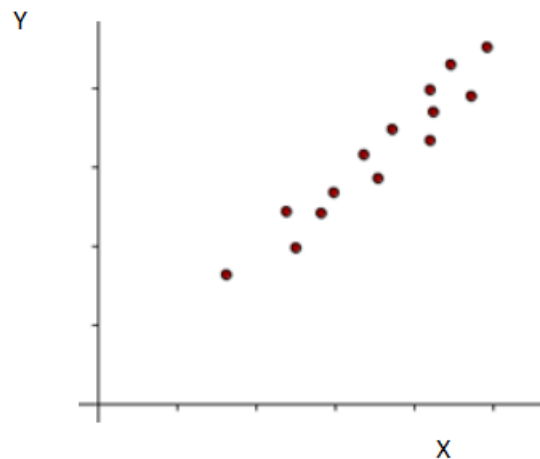
iii) Si dos variables X e Y son independientes entonces:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = 0.$$

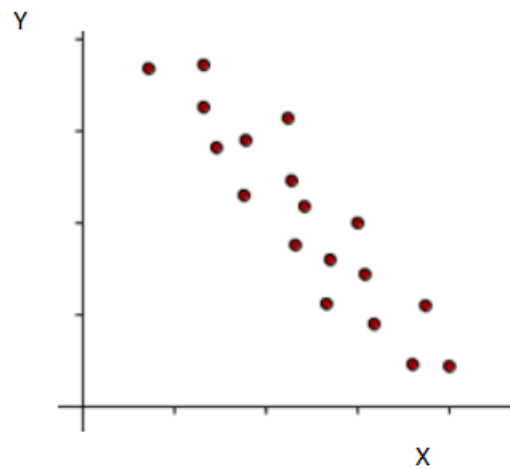
Sin embargo el recíproco no es cierto. Como ya hemos visto en el ejemplo de la suma y diferencia de puntos al lanzar dos dados.

Interpretación del coeficiente de correlación lineal

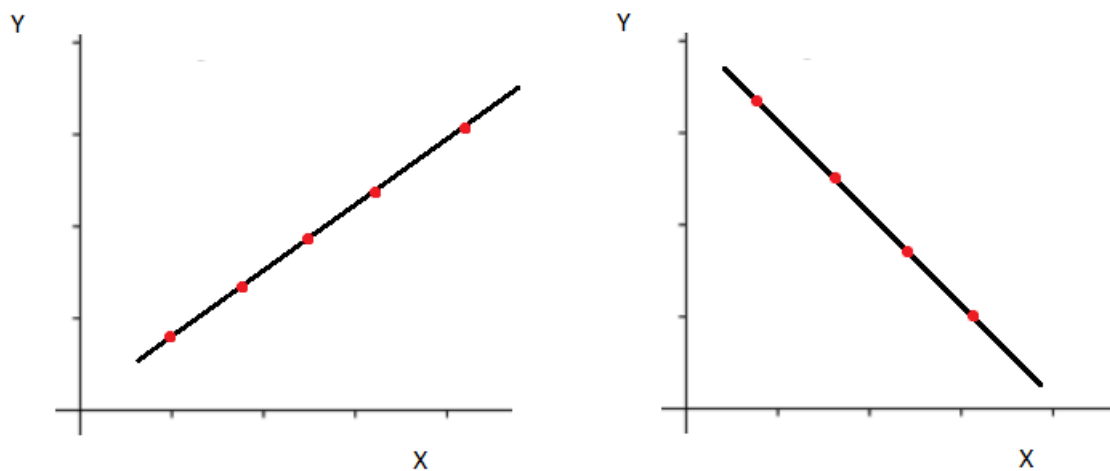
El coeficiente de correlación lineal se interpreta como el grado de dependencia lineal existente entre dos variables aleatorias. Cuanto más grande sea su valor en módulo mas grande será la dependencia lineal. El signo solo afecta al tipo de relación entre las variables. El signo positivo indica que cuando una variable incrementa su valor, también lo hace la otra. Por ejemplo la renta y el consumo.



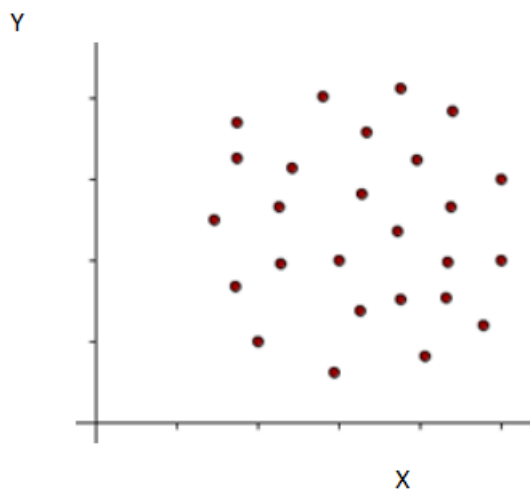
El signo negativo indica que cuando una variable incrementa su valor la otra disminuye. Por ejemplo, tipos de interés y consumo.



Si $\rho = \pm 1$, X e Y presentan dependencia lineal exacta. Una variable puede ser exactamente pronosticada a partir de la otra. Ejemplo, la fuerza y la aceleración para una masa fija, $F = m \cdot a$, según la ley de Newton. Pero este tipo de situaciones frontera no interesan en estadística porque realmente no corresponden a fenómenos aleatorios, sino a fenómenos deterministas.



Por el contrario, si $\rho = 0$ el grado de dependencia lineal es nula. Y una de las variables no se podrá predecir ni siquiera de forma aproximada, a partir de la otra.



5.4 Cambio de variable

Dado un vector aleatorio bidimensional (X, Y) con densidad estrictamente positiva en $S \in \mathbb{R}^2$ con distribución conocida, se obtiene un nuevo vector aleatorio bidimensional (U, V) mediante las transformaciones: $U = g_1(X, Y)$ y $V = g_2(X, Y)$, de tal forma que $(U, V) \in T$ si y solo si $(X, Y) \in S$.

Caso discreto

Sea (X, Y) un vector aleatorio discreto con función de probabilidad conjunta $P(X = i, Y = j)$, $\forall i \in I$ y $\forall j \in J$. Sea un nuevo vector $(U, V) = (g_1(X, Y), g_2(X, Y))$. Entonces la función de probabilidad de (U, V) vendrá dada por:

$$P(U = u, V = v) = \sum_{(i,j)/g_1(i,j)=u, g_2(i,j)=v} P(X = i, Y = j)$$

Ejemplo 5.8:

Sea (X, Y) el vector aleatorio mínimo y máximo de puntos al lanzar dos dados, y sea: $U = X/Y$ y $V = |X - Y|$.

La función de probabilidad de (U, V) será:

| U/V | 0 | 1 | 2 | 3 | 4 | 5 | U |
|-----|------|-------|------|------|------|------|------|
| 1 | 6/36 | 0 | 0 | 0 | 0 | 0 | 6/36 |
| 1/2 | 0 | 2/36 | 2/36 | 2/36 | 0 | 0 | 6/36 |
| 1/3 | 0 | 0 | 2/36 | 0 | 2/36 | 0 | 4/36 |
| 1/4 | 0 | 0 | 0 | 2/36 | 0 | 0 | 2/36 |
| 1/5 | 0 | 0 | 0 | 0 | 2/36 | 0 | 2/36 |
| 1/6 | 0 | 0 | 0 | 0 | 0 | 2/36 | 2/36 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2/3 | 0 | 2/36 | 2/36 | 0 | 0 | 0 | 4/36 |
| 2/5 | 0 | 0 | 0 | 2/36 | 0 | 0 | 2/36 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3/2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3/4 | 0 | 2/36 | 0 | 0 | 0 | 0 | 2/36 |
| 3/5 | 0 | 0 | 2/36 | 0 | 0 | 0 | 2/36 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4/3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4/5 | 0 | 2/36 | 0 | 0 | 0 | 0 | 2/36 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5/2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5/3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5/4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5/6 | 0 | 2/36 | 0 | 0 | 0 | 0 | 2/36 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6/5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| V | 6/36 | 10/36 | 8/36 | 6/36 | 4/36 | 2/36 | 1 |

Caso absolutamente continuo

Teorema Jacobiano Sea (X, Y) un vector aleatorio absolutamente continuo con densidad $f(x, y)$ estrictamente positiva en el recinto $S \subseteq \mathbb{R}^2$. Sea (U, V) una transformación continua y biyectiva de (X, Y) . Entonces la densidad de (U, V) es:

$$g(u, v) = \begin{cases} f_{(x,y)}(x(u, v), y(u, v)) |J| & \text{si } (u, v) \in T \\ 0 & \text{fuera.} \end{cases}$$

donde $(x(u, v), y(u, v))$ es la transformación inversa (escribir x, y en función de (u, v))

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix},$$

y T es el recinto transformado de S biyectivamente.

Ejemplo 5.9:

Sea (X, Y) el vector aleatorio bidimensional con densidad uniforme en el cuadrado $(0, 0)$, $(0, 1)$,

$(1,0), (1,1)$. Sea $U = \frac{X}{Y}$ y $V = X \cdot Y$. Calcular la densidad de (U, V) y las marginales de U y V .

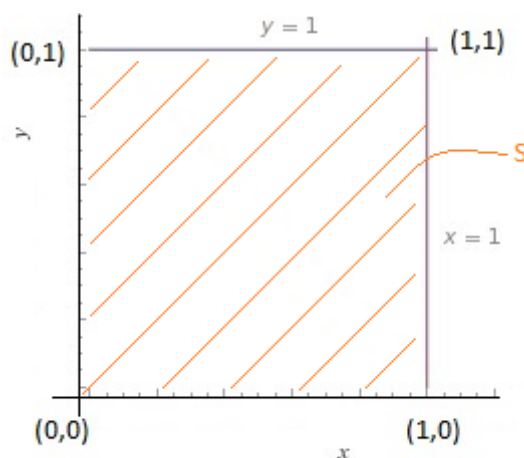


Figura 5.1: Recinto S

$$f(x,y) = \begin{cases} 1 & \text{si } (x,y) \in S \\ 0 & \text{fuera} \end{cases}$$

transformación directa:

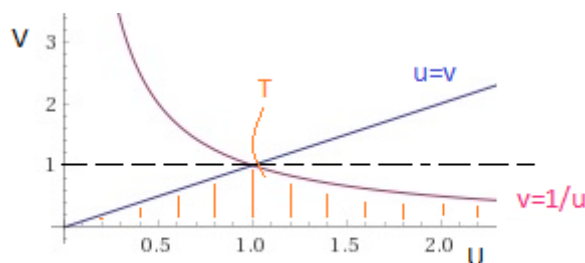
$$u = x/y, v = x \cdot y$$

transformación inversa:

$$x = \sqrt{v \cdot u}, y = \sqrt{v/u}$$

Restricciones del recinto T:

$$0 \leq x \leq 1, 0 \leq y \leq 1 \Rightarrow 0 \leq \sqrt{v \cdot u} \leq 1, 0 \leq \sqrt{v/u} \leq 1 \Rightarrow 0 \leq u, 0 \leq v \leq 1, v \leq \frac{1}{u}, v \leq u$$



Densidad conjunta:

$$J = \left| \begin{array}{cc} \frac{\sqrt{v}}{2\sqrt{u}} & \frac{\sqrt{u}}{2\sqrt{v}} \\ \sqrt{v}(\frac{-1}{2})u^{-3/2} & \frac{1}{2\sqrt{uv}} \end{array} \right| = \frac{1}{4u} + \frac{1}{4u} = \frac{1}{2u}$$

$$f(u, v) = \begin{cases} 1 \cdot \left| \frac{1}{2u} \right| = \frac{1}{2u} & \text{si } (u, v) \in T \\ 0 & \text{fuera.} \end{cases}$$

Marginales:

$$f(u) = \begin{cases} \int_0^u \frac{1}{2u} dv = \frac{1}{2} & \text{si } u \in [0, 1] \\ \int_0^{1/u} \frac{1}{2u} dv = \frac{1}{2u^2} & \text{si } u \geq 1 \\ 0 & \text{fuera.} \end{cases}$$

$$f(v) = \begin{cases} \int_v^{1/v} \frac{1}{2u} du = \frac{1}{2}(-\ln v - \ln v) = -\ln v & \text{si } v \in (0, 1) \\ 0 & \text{fuera.} \end{cases}$$

Ejemplo 5.10:

Sea (X, Y) el vector aleatorio bidimensional con densidad uniforme en el cuadrado $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$. Calcular la densidad de $U = X + Y$.

Aquí la transformación no preserva la dimensión, pues estamos pasando de dos variables X e Y a una sola U . Para poder aplicar el teorema jacobiano necesitamos utilizar una variable auxiliar (que elegiremos lo más sencillamente posible). Cuando hayamos hecho la transformación de (X, Y) a (U, V) el problema quedará resuelto calculando la marginal de U .

Utilizamos la variable auxiliar $V = X$, por ejemplo.

$$f(x, y) = \begin{cases} 1 & \text{si } (x, y) \in S \\ 0 & \text{fuera.} \end{cases}$$

transformaciones directas:

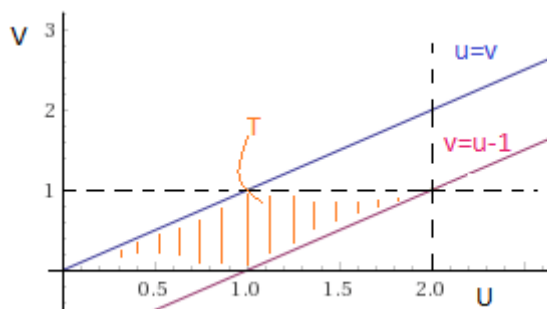
$$u = x + y, v = x$$

transformaciones inversas:

$$x = v, y = u - v$$

Restricciones del recinto T:

$$0 \leq x \leq 1, 0 \leq y \leq 1 \Rightarrow 0 \leq u \leq 2, 0 \leq v \leq 1, 0 \leq u - v \leq 1 \Rightarrow u \geq v, v \geq u - 1$$



Densidad conjunta:

$$J = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1$$

$$f(u, v) = \begin{cases} 1 \cdot |-1| = 1 & \text{si } (u, v) \in T \\ 0 & \text{fuera.} \end{cases}$$

Marginal de U:

$$f(u) = \begin{cases} \int_0^u 1 \, dv = u & \text{si } u \in (0, 1) \\ \int_{u-1}^1 1 \, dv = 2 - u & \text{si } u \in (1, 2) \\ 0 & \text{fuera.} \end{cases}$$

Ejemplo 5.11:

Sea (X, Y) el vector aleatorio bidimensional con variables independientes tales que:

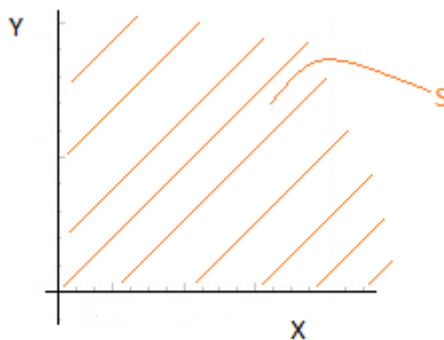
$$f(x) = \begin{cases} e^{-x} & \text{si } x \geq 0 \\ 0 & \text{fuera.} \end{cases}$$

$$f(y) = \begin{cases} e^{-y} & \text{si } y \geq 0 \\ 0 & \text{fuera.} \end{cases}$$

Sea $U = X + Y$ y $V = X - Y$. Calcular la densidad conjunta (U, V) y sus marginales.

Como las variables son X e Y son independientes, entonces se puede calcular la función de densidad conjunta como el producto de las densidades marginales:

$$f(x, y) = f(x) \cdot f(y) = \begin{cases} e^{-(x+y)} & \text{si } x \geq 0, y \geq 0 \\ 0 & \text{fuera.} \end{cases}$$



Transformaciones directas:

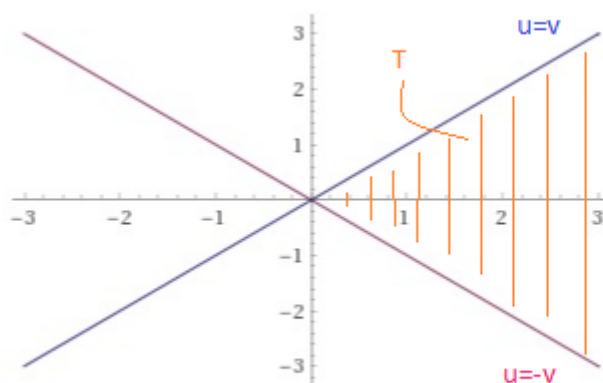
$$u = x + y, v = x - y$$

Transformaciones inversas:

$$x = \frac{u + v}{2}, y = \frac{u - v}{2}$$

Restricciones del recinto T:

$$0 \leq x, 0 \leq y \Rightarrow 0 \leq \frac{u+v}{2}, 0 \leq \frac{u+v}{2}, u \geq 0 \Rightarrow 0 \leq u, u \geq -v, u \geq v$$

**Densidad conjunta:**

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}$$

$$f(u, v) = \begin{cases} f_{(x,y)}(x(u, v), y(u, v)) |J| = e^{-u} \cdot \left| -\frac{1}{2} \right| = \frac{1}{2} e^{-u} & \text{si } (u, v) \in T \\ 0 & \text{fuera.} \end{cases}$$

Marginales:

$$f(u) = \begin{cases} \int_{-u}^u e^{-u} \cdot \frac{1}{2} dv = u e^{-u} & \text{si } u \geq 0 \\ 0 & \text{fuera.} \end{cases}$$

$$f(v) = \begin{cases} \int_v^\infty e^{-u} \cdot \frac{1}{2} du = \frac{1}{2} e^{-v} & \text{si } v \geq 0 \\ \int_{-v}^\infty e^{-u} \cdot \frac{1}{2} du = \frac{1}{2} e^v & \text{si } v < 0 \\ 0 & \text{fuera.} \end{cases}$$

Ejemplo 5.12:

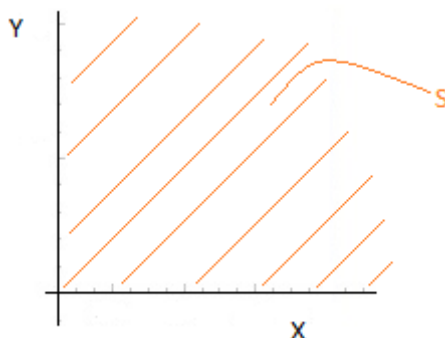
Sean X e Y variables aleatorias independientes, ambas con distribución $\exp(1)$. Calcular la densidad conjunta y las densidades marginales de $U = \frac{X}{X+Y}$ y $V = X+Y$.

$$f(x) = \begin{cases} e^{-x} & \text{si } x \geq 0 \\ 0 & \text{fuera} \end{cases}$$

$$f(y) = \begin{cases} e^{-y} & \text{si } y \geq 0 \\ 0 & \text{fuera} \end{cases}$$

Como las variables son X e Y son independientes:

$$f(x,y) = f(x) \cdot f(y) = \begin{cases} e^{-(x+y)} & \text{si } x \geq 0, y \geq 0 \\ 0 & \text{fuera} \end{cases}$$



transformaciones directas:

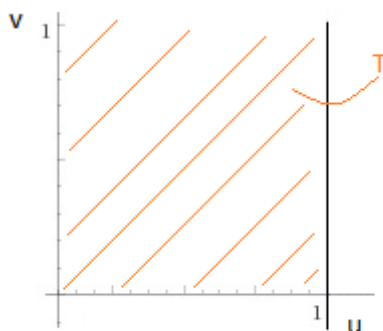
$$u = \frac{x}{x+y}, v = x+y$$

transformaciones inversas:

$$x = u \cdot v, y = v - uv$$

Restricciones del recinto T:

$$0 \leq x, 0 \leq y \Rightarrow 0 \leq u \leq 1, 0 \leq v$$



Densidad conjunta:

$$J = \begin{vmatrix} v & u \\ -v & 1-u \end{vmatrix} = v(1-u) + uv = v$$

$$f(u, v) = \begin{cases} e^{-v} \cdot |v| = v \cdot e^{-v} & \text{si } (u, v) \in T \\ 0 & \text{fuera.} \end{cases}$$

Marginales:

$$\begin{aligned} f(u) &= \int_0^{+\infty} v \cdot e^{-v} dv = \int_0^{+\infty} x \cdot e^{-x} dx \\ &\stackrel{dv=e^{-x} dx, v=e^{-x}}{u=x, du=dx} -xe^{-x} + \int_0^{+\infty} e^{-x} dx = \begin{cases} 1 & \text{si } u \in [0, 1] \\ 0 & \text{fuera.} \end{cases} \\ U &\sim U(0, 1). \end{aligned}$$

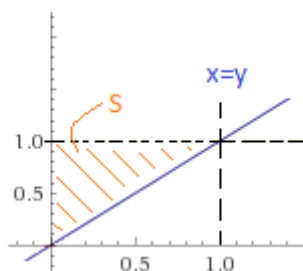
$$f(v) = \begin{cases} \int_0^1 v \cdot e^{-v} du = v \cdot e^{-v} & \text{si } v \geq 0 \\ 0 & \text{fuera} \end{cases} \sim \Gamma(2, 1).$$

Ejemplo 5.13:

Sean X e Y variables aleatorias con función de densidad conjunta

$$f(x, y) = \begin{cases} 2(x+y) & \text{si } 0 \leq x \leq y \leq 1 \\ 0 & \text{fuera.} \end{cases}$$

Calcular la densidad de $U = X + Y$, eligiendo $V = X$ como variable auxiliar.



transformaciones directas:

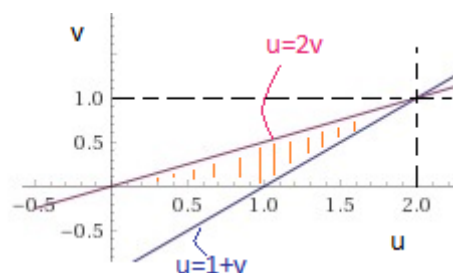
$$u = x + y, v = x$$

transformaciones inversas:

$$x = v, y = u - v$$

Restricciones del recinto T:

$$0 \leq x \leq y \leq 1 \Rightarrow v \leq u - v, 0 \leq u - v \leq 1 \Rightarrow 0 \leq u \leq 2, u \leq 1 + v, u \geq 2v.$$



Densidad conjunta:

$$J = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1$$

$$f(u, v) = \begin{cases} 2u \cdot |-1| = 2u & \text{si } (u, v) \in T \\ 0 & \text{fuera.} \end{cases}$$

Marginales:

$$f(u) = \begin{cases} \int_0^{\frac{u}{2}} 2u \, dv = u^2 & \text{si } u \in (0, 1) \\ \int_{u-1}^{\frac{u}{2}} 2u \, dv = (2-u)u & \text{si } u \in (1, 2) \\ 0 & \text{fuera.} \end{cases}$$

$$f(v) = \begin{cases} \int_{2v}^{1+v} 2u \, du = 1 - 2v - 3v^2 & \text{si } v \in (0, 1) \\ 0 & \text{fuera.} \end{cases}$$

Ejemplo 5.14:

Repetición del ejercicio anterior eligiendo esta vez $V = Y$ como variable auxiliar.

transformación directa:

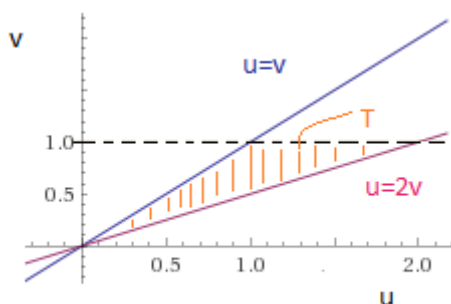
$$U = X + Y, V = Y$$

transformación inversa:

$$Y = V, X = U - V$$

Restricciones del recinto T:

$$0 \leq u \leq 2, u - v \leq v \rightarrow u \leq 2v, 0 \leq v \leq 1, 0 \leq u - v \leq 1 \rightarrow u \leq 1 + v, u \geq v.$$



Densidad conjunta:

$$J = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$$

$$f(u, v) = \begin{cases} 2(u - v + v)|1| = 2u & \text{si } (u, v) \in T \\ 0 & \text{fuera.} \end{cases}$$

Densidad de U:

$$f(u) = \begin{cases} \int_{\frac{u}{2}}^u 2u \, dv = u^2 & \text{si } u \in (0, 1) \\ \int_{\frac{u}{2}}^1 2u \, dv = u(2 - u) & \text{si } u \in (1, 2) \\ 0 & \text{fuera.} \end{cases}$$

Cambio de variable en el caso N-dimsional

En el caso continuo y para transformaciones biyectivas, se dispone del teorema jacobiano n-dimensional.

Teorema Jacobiano Sea (X_1, \dots, X_n) un vector aleatorio n-dimensional absolutamente continuo con densidad $f(x_1, \dots, x_n) > 0$ si $(x_1, \dots, x_n) \in S$. Y sea $g(X_1, \dots, X_n) = (Y_1, \dots, Y_n)$ una transformación continua y biyectiva de (X_1, \dots, X_n) en (Y_1, \dots, Y_n) que aplica S en T . Entonces,

$$\begin{aligned} & f_{(Y_1, \dots, Y_n)}(y_1, \dots, y_n) \\ &= \begin{cases} f_{(X_1, \dots, X_n)}(x_1(y_1, \dots, y_n), \dots, x_n(y_1, \dots, y_n)) \cdot |J| & \text{si } (y_1, \dots, y_n) \in T \\ 0 & \text{fuera.} \end{cases} \end{aligned}$$

En general, el teorema Jacobiano en el caso n-dimensional es de difícil aplicabilidad. En determinadas situaciones, sin embargo, es muy útil conocer la distribución de ciertas variables de un vector n-dimensional y no siempre se utiliza el teorema Jacobiano, a veces, por falta de biyectividad o por tratarse de casos discretos.

Ejemplo 5.15:

Sea (X_1, X_2, \dots, X_n) un vector aleatorio n-dimensional de variables independientes tales que cada variable $X_i \sim B(1, p)$ (sigue una distribución de Bernoulli), $i = 1, \dots, n$. Consideramos las nuevas variables $Y = \max\{X_1, X_2, \dots, X_n\}$, $Z = \min\{X_1, X_2, \dots, X_n\}$. Calcular distribución de Y y de Z .

Distribución de Y:

$$P(Y = 0) = \prod_{i=1}^n P(X_i = 0) = (1 - p)^n, \quad P(Y = 1) = 1 - (1 - p)^n.$$

Distribución de Z.

$$P(Z = 1) = \prod_{i=1}^n P(X_i = 1) = p^n, \quad P(Z = 0) = 1 - p^n.$$

Ejemplo 5.16:

Consideramos un sistema electrónico formado por 4 componentes que funcionan independientemente unos de otros. Sea T_i la variable aleatoria “tiempo de vida del i -ésimo componente” para $i = 1, 2, 3, 4$. Supongamos que los componentes son iguales y que por tanto las cuatro variables están igualmente distribuidas con función de distribución $F(t)$ y función de densidad $f(t)$. Sea T la variable aleatoria “tiempo de vida del sistema”. Determinar la función de distribución de T y su función de densidad si:

- a) El sistema está montado en serie.
b) El sistema está montado en paralelo.

Aplicar al caso particular de distribuciones exponenciales para el caso de los tiempos de vida de los componentes.

- a) Sea T el tiempo de vida del sistema montado en serie (falla el sistema cuando falla un componente). $T = \min\{T_1, T_2, T_3, T_4\}$,

$$\begin{aligned} F_T(t) &= P(T \leq t) = P(\min_{i=1,2,3,4} \{T_i\} \leq t) = 1 - P(\min_{i=1,2,3,4} \{T_i\} \geq t) \\ &= 1 - P(T_1 > t, T_2 > t, T_3 > t, T_4 > t) = 1 - \prod_{i=1}^4 P(T_i \geq t) \\ &= 1 - \prod_{i=1}^4 [1 - P(T_i \leq t)] = 1 - [1 - F(t)]^4 \end{aligned}$$

$f_T(t) = F'_T(t) = 4[1 - F(t)]^3 \cdot f(t)$. Si $T_i \sim \exp(\lambda)$, $i = 1, 2, 3, 4$, entonces,

$$F(t) = \int_0^t \lambda e^{-\lambda s} ds = \begin{cases} 1 - e^{-\lambda t} & \text{si } t \geq 0 \\ 0 & \text{fuera.} \end{cases}$$

$$f_T(t) = \begin{cases} 4(e^{-\lambda t})^3 \lambda e^{-\lambda t} & \text{si } t \geq 0 \\ 0 & \text{fuera} \end{cases} = \begin{cases} 4\lambda e^{-4\lambda t} & \text{si } t \geq 0 \\ 0 & \text{fuera.} \end{cases} \sim \exp(4\lambda)$$

- b) Sea U el tiempo de vida del sistema montado en serie (falla el sistema cuando falla el último componente). $U = \max_{i=1,2,3,4} \{T_i\}$

$$\begin{aligned} F_U(u) &= P(U \leq u) = P(\max_{i=1,2,3,4} \{T_i\} \leq u) = P(T_1 \leq u, T_2 \leq u, T_3 \leq u, T_4 \leq u) \\ &= \prod_{i=1}^4 P(T_i \leq u) = [F(t)]^4 \end{aligned}$$

$f_U(u) = F'_U(u) = 4[F(u)]^3 f(u)$. Si $T_i \sim \exp(\lambda)$, $i = 1, 2, 3, 4$,

$$f_U(u) = \begin{cases} 4(1 - e^{-\lambda u})^3 \lambda e^{-\lambda u} & \text{si } u \geq 0 \\ 0 & \text{fuera.} \end{cases}$$

Bibliografía

1. A.A. Borovkov (2013). *Probability Theory*. Springer.
2. A. Klenke (2014). *Probability Theory. A Comprehensive Course*. Springer.
3. M.H. DeGroot (1988). *Probability and Statistics*. Wilmington, Delaware, E.U.A. Addison-Wesley Iberoamericana, S.A.
4. V.K. Rohatgi and A.K. Md. Ehsanes Saleh (2001). *An Introduction to Probability and Statistics*. John Wiley and Sons.

Tema 6

Propiedades de una muestra aleatoria. Conceptos básicos de muestras aleatorias. Sumas de variables aleatorias de una muestra aleatoria. Muestreo de una distribución normal: propiedades de la media y varianza muestrales y las distribuciones t de Student y F de Snedecor. Estadísticos de orden.

6.1 Modelo paramétrico y modelo no paramétrico

En inferencia estadística llamaremos modelo paramétrico a $\{f(x, \theta), \theta \in \Theta\}$, donde $f(x, \theta)$ es la función de probabilidad o de densidad de una variable aleatoria X , la cual, en este contexto inferencial recibe el nombre de población, θ representa un parámetro o vector de parámetros desconocido(s) hacia el que se dirigen nuestras inferencias y Θ es el conjunto de valores posibles para θ y recibe el nombre de espacio paramétrico.

Ejemplo 6.1: La intención de voto en un parlamento con dos partidos se estudia a través del modelo paramétrico.

$$X \sim B(1, p), p \in (0, 1)$$

Ejemplo 6.2: Para saber la cantidad de dependientes que necesita tener el Corte Inglés en la campaña navideña se utiliza el modelo.

$$X \sim P(\lambda), \lambda \in \mathbb{R}^+.$$

Este modelo puede servir también para los boxes abiertos en un peaje en cada franja horaria, cajas abiertas en un supermercado en diferentes horarios...

Ejemplo 6.3: Estimación del consumo de los hogares.

$$X \sim N(\mu, \sigma), \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+.$$

Si el modelo es no paramétrico, entonces los problemas son otros:

- a) Decidir qué tipo de población subyace en los datos muestreados.
- b) ¿La muestra es aleatoria?
- c) En general cuando se dispone de pocas observaciones o la escala de medición es baja, entonces se utilizan técnicas no paramétricas.

Situados en el modelo paramétrico, la inferencia sobre el parámetro desconocido θ descansa sobre 3 técnicas:

- a) Estimación puntual: Consiste en hacer un pronóstico (técnicamente una estimación) sobre el valor del parámetro desconocido a partir de un conjunto de observaciones.

b) Intervalos de confianza: Es un rango de valores en el que existe una probabilidad alta, normalmente del 95 % de encontrar el valor del parámetro desconocido.

c) Contraste de hipótesis: Se trata de una regla de decisión para elegir entre dos conjuntos de valores posibles para el parámetro desconocido. Ejemplo, en $X \sim B(1, p)$.

$$p < \frac{1}{2} \text{ frente a } p \geq \frac{1}{2}.$$

6.2 Conceptos básicos sobre inferencia.

A continuación se definen varios conceptos esenciales para la realización de cualquier tipo de inferencia.

i) Población: Colectivo sujeto del estudio. Lo identificaremos con la magnitud que se desea estudiar y que vendrá dada por una variable aleatoria.

Así, por ejemplo, si queremos analizar las estaturas de los españoles, la población sería el conjunto de todas las estaturas de todos los españoles, siendo el universo el conjunto de todos los españoles.

ii) Muestra: Un subconjunto cualquiera de la población. Para que la muestra nos sirva para extraer conclusiones sobre la población deber ser representativa, lo que se consigue seleccionando sus elementos al azar, lo que da lugar a una muestra aleatoria.

iii) Muestreo: Procedimiento para seleccionar individuos en la población. Se requiere que la muestra refleje lo más fielmente posible a la población de la que es extraída. Existen diferentes tipos.

a) Muestreo aleatorio: Es aquel procedimiento de selección de la muestra en el que todos y cada uno de los elementos de la población tiene una cierta probabilidad de resultar elegidos. De esta forma, si tenemos una población de N elementos y estamos interesados en obtener una muestra de n elementos (muestra de tamaño n), cada subconjunto de n elementos de la población tendrá también una cierta probabilidad de resultar la muestra elegida.

b) Muestreo aleatorio simple: Es aquel en el que todos los individuos de la población tienen la misma probabilidad de ser elegidos. Matemáticamente, si X es la variable aleatoria población, una muestra aleatoria simple (m.a.s.) de tamaño n , se formula por n variables aleatorias independientes idénticamente distribuidas, X_1, X_2, \dots, X_n , todas ellas con la misma distribución que X . X_i representa la aleatoriedad del i -ésimo individuo elegible en la muestra.

c) Muestreo errático/sin norma: Los encuestados van entrando sin ningún tipo de criterio y no son representativos de la población global. Es frecuente en programas de televisión o radio

d) Muestreo por bola de nieve: Se dirige a un grupo concreto de la población. Se busca que la muestra esté correlacionada precisamente porque todos los individuos pertenecen al sector. Muy

utilizado en venta a domicilio o en publicidad en internet.

iv) Estadístico: Es cualquier función de los valores muestrales que dependa exclusivamente de estos. En la medida en que los valores muestrales son variables aleatorias también lo serán las funciones de éstos: los estadísticos. Dado que los estadísticos son variables aleatorias, tendrán determinadas distribuciones de probabilidad y determinados parámetros (media, varianza, etc). Para el desarrollo de la inferencia es imprescindible conocer dichas distribuciones y parámetros, consiguiendo establecer entonces las relaciones entre éstas y las de la población, pudiendo entonces inferir las características desconocidas de dicha población.

6.3 Estadísticos muestrales. Propiedades

Definición 6.1: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de X . Llamaremos estadístico muestral o simplemente estadístico a cualquier función de la muestra y sólo de la muestra, es decir, no debe depender de parámetros desconocidos. Cuando los estadísticos se utilizan en problemas de estimación reciben el nombre de estimadores. Algunos de amplio uso son:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ (Media muestral).}$$

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ (Varianza muestral).}$$

$$\hat{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ (Covarianza muestral).}$$

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \text{ (Momento muestral de orden k).}$$

$$S = \sqrt{S^2} \text{ (Desviación típica muestral).}$$

$$\hat{S} = \sqrt{\hat{S}^2} \text{ (Cuasidesviación típica muestral).}$$

Todos los estadísticos son variables aleatorias dado que son funciones de la muestra que está integrada por variables aleatorias. Entonces no deben confundirse con los parámetros poblacionales $\mu = EX$, $\sigma^2 = VarX$, $\sigma = \sqrt{VarX}$, EX^k . Para enfatizar esta distinción, en lo que sigue, a μ le denominaremos media poblacional, a σ^2 la varianza poblacional, a σ la desviación típica poblacional y a EX^k momento poblacional de orden k.

Es un grave error escribir o confundir $\bar{X} = \mu$, $S^2 = \sigma^2$, $\hat{S}^2 = \sigma^2$ o $A_k = EX^k$.

A su vez, asumiendo que \bar{X} , S^2 , \hat{S}^2 o A_k son variables aleatorias tendrá sentido preguntarnos por $E\bar{X}$, $Var\bar{X}$, ES^2 , $VarS^2$, etc.

Sin embargo, no es menos cierto que los estadísticos muestrales, en problemas de estimación se utilizan como estimadores de los parámetros poblacionales. Es decir, utilizamos en general, \bar{X} para estimar μ ; S^2 para estimar σ^2 ; S y \hat{S} para estimar σ y finalmente A_k para estimar EX^k .

Proposición 6.1: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de una población X , variable aleatoria, tal que $EX = \mu$ y $VarX = \sigma^2$. Se tiene:

- a) $E\bar{X} = \mu$, es decir, la esperanza de la media muestral es la media poblacional.
- b) $Var\bar{X} = \frac{\sigma^2}{n}$, es decir, la varianza de la media muestral es la varianza poblacional dividida entre el tamaño muestral.
- c) $E(S^2) = \frac{n-1}{n}\sigma^2$. La esperanza de la varianza muestral es menor que la varianza poblacional.
- d) $E(\hat{S}^2) = \sigma^2$. La esperanza de la cuasivarianza muestral es la varianza poblacional.

Demostración:

a)

$$E\bar{X} = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n EX_i = \frac{1}{n} \sum_{i=1}^n EX = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu.$$

b)

$$\begin{aligned} Var\bar{X} &= Var\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{1}{n^2} \sum_{i=1}^n Var(X) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2 n}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

c) Probemos, en primer lugar que $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2$. En efecto,

$$\begin{aligned} S^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n (X_i^2 + \bar{X}^2 - 2X_i\bar{X}) = \frac{1}{n} \sum_{i=1}^n X_i^2 + \bar{X}^2 - 2\bar{X}\bar{X} \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \end{aligned}$$

Entonces,

$$\begin{aligned} E(S^2) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2\right) = \frac{1}{n} \sum_{i=1}^n EX_i^2 - E\bar{X}^2 = \frac{1}{n} \sum_{i=1}^n EX^2 - E\bar{X}^2 \\ &=_{VarX=EX^2-E^2X} \frac{1}{n} n(\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) = \frac{n-1}{n}\sigma^2 \end{aligned}$$

d)

$$E(\hat{S}^2) = E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) = E\left(\frac{n}{n-1} S^2\right) = \frac{n}{n-1} ES^2 = \frac{n}{n-1} \sigma^2 \frac{n-1}{n} = \sigma^2$$

Como consecuencia de los resultados anteriores:

a) Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $B(1, p)$ entonces $E\bar{X} = E(B(1, p)) = p$, $Var(\bar{X}) = \frac{Var(B(1, p))}{n} = \frac{pq}{n}$, $ES^2 = \frac{n-1}{n} Var(B(1, p)) = \frac{n-1}{n} pq$ y $E\hat{S}^2 = Var(B(1, p)) = pq$.

b) Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $Bin(m, p)$ entonces $E\bar{X} = E(Bin(m, p)) = mp$, $Var(\bar{X}) = \frac{Var(Bin(m, p))}{n} = \frac{mpq}{n}$, $ES^2 = \frac{n-1}{n} Var(Bin(m, p)) = \frac{n-1}{n} mpq$ y $E\hat{S}^2 = Var(Bin(m, p)) = mpq$.

c) Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $P(\lambda)$ entonces $E\bar{X} = E(P(\lambda)) = \lambda$, $Var(\bar{X}) = \frac{Var(P(\lambda))}{n} = \frac{\lambda}{n}$, $ES^2 = \frac{n-1}{n} Var(P(\lambda)) = \frac{n-1}{n} \lambda$ y $E\hat{S}^2 = Var(P(\lambda)) = \lambda$.

d) Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $U(a, b)$ entonces $E\bar{X} = E(U(a, b)) = \frac{a+b}{2}$, $Var(\bar{X}) = \frac{Var(U(a, b))}{n} = \frac{(b-a)^2}{12n}$, $ES^2 = \frac{n-1}{n} Var(U(a, b)) = \frac{(n-1)(b-a)^2}{12n}$ y $E\hat{S}^2 = Var(U(a, b)) = \frac{(b-a)^2}{12}$.

e) Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $\Gamma(a, p)$ entonces $E\bar{X} = E(\Gamma(a, p)) = \frac{a}{p}$, $Var(\bar{X}) = \frac{Var(\Gamma(a, p))}{n} = \frac{a}{np^2}$, $ES^2 = \frac{n-1}{n} Var(\Gamma(a, p)) = \frac{n-1}{n} \frac{a}{p^2}$ y $E\hat{S}^2 = Var(\Gamma(a, p)) = \frac{a}{p^2}$.

f) Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $N(\mu, \sigma)$ entonces $E\bar{X} = E(N(\mu, \sigma)) = \mu$, $Var(\bar{X}) = \frac{Var(N(\mu, \sigma))}{n} = \frac{\sigma^2}{n}$, $ES^2 = \frac{n-1}{n} Var(N(\mu, \sigma)) = \frac{n-1}{n} \sigma^2$ y $E\hat{S}^2 = Var(N(\mu, \sigma)) = \sigma^2$.

Ejemplo 6.4: Hallar la distribución exacta de la media muestral si la muestra procede de una población X tal que:

- a) $X \rightsquigarrow B(1, p)$.
- b) $X \rightsquigarrow \exp(\lambda)$.
- c) $X \rightsquigarrow Cauchy$.
- d) $X \rightsquigarrow N(\mu, \sigma)$.

a) \bar{X} se interpreta como la frecuencia relativa de éxitos.

$$\begin{aligned}\varphi_{\bar{X}}(t) &= E(e^{it\bar{X}}) = E(e^{i\frac{t}{n} \sum_{j=1}^n X_j}) = E\left(\prod_{j=1}^n e^{i\frac{tX_j}{n}}\right) = \prod_{j=1}^n E(e^{i\frac{tX_j}{n}}) = \prod_{j=1}^n \varphi_{X_j}\left(\frac{t}{n}\right) \\ &= \prod_{j=1}^n [pe^{i\frac{t}{n}} + (1-p)] = [pe^{i\frac{t}{n}} + (1-p)]^n.\end{aligned}$$

No sigue ninguna distribución teórica conocida.

b)

$$\varphi_{\bar{X}}(t) = E(e^{it\bar{X}}) = \prod_{j=1}^n \varphi_{X_j}\left(\frac{t}{n}\right) = \left(\frac{\lambda}{\lambda - i\frac{t}{n}}\right)^n = \left(\frac{n\lambda}{n\lambda - it}\right)^n \rightsquigarrow \Gamma(n, n\lambda)$$

c)

$$\varphi_{\bar{X}}(t) = E(e^{it\bar{X}}) = \prod_{j=1}^n \varphi_{X_j}\left(\frac{t}{n}\right) = \prod_{j=1}^n e^{-|\frac{t}{n}|} = e^{-|\frac{t}{n}|n} = e^{-|t|} \rightsquigarrow \text{Cauchy}$$

d)

$$\varphi_{\bar{X}}(t) = E(e^{it\bar{X}}) = \prod_{j=1}^n \varphi_{X_j}\left(\frac{t}{n}\right) = \left(e^{i\frac{\mu}{n} - \frac{t^2\sigma^2}{2n^2}}\right)^n = e^{it\mu - \frac{1}{2}\left(\frac{\sigma}{\sqrt{n}}\right)^2 t^2} \rightsquigarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Ejemplo 6.5: En una urna hay 100 bolas numeradas. Se extraen 10 bolas con remplazamiento. Sea \bar{X} el estadístico media muestral de los números obtenidos. Determinar $E\bar{X}$ y $Var\bar{X}$.

Disponemos de una muestra aleatoria simple X_1, X_2, \dots, X_{10} de una población X tal que $P(X = k) = \frac{1}{100}$, $k = 1, 2, \dots, 100$. Sabemos que $E\bar{X} = EX$ y $Var\bar{X} = \frac{VarX}{n}$.

$$E\bar{X} = EX = \sum_{i=1}^{100} k\left(\frac{1}{100}\right) = \frac{1}{100} \sum_{i=1}^{100} k = \frac{1}{100} \left(\frac{100 \cdot 101}{2}\right) = \frac{101}{2},$$

$$EX^2 = \sum_{i=1}^{100} k^2\left(\frac{1}{100}\right) = \frac{1}{100} \sum_{i=1}^{100} k^2 = \frac{1}{100} \left(\frac{100 \cdot 101 \cdot 201}{6}\right) = \frac{6767}{2},$$

$$Var\bar{X} = \frac{VarX}{10} = \frac{1}{10}(EX^2 - E^2X) = 83,32.$$

Ejemplo 6.6: Sea X una población con función de probabilidad:

| | | | |
|------------|---------------|---------------|---------------|
| X | 0 | 1 | 2 |
| $P(X = x)$ | $\frac{1}{6}$ | $\frac{1}{2}$ | $\frac{1}{3}$ |

Se selecciona una muestra de tamaño $n = 2$ y se pide:

- Obtener la función de probabilidad de la muestra aleatoria simple
- Obtener la función de probabilidad de \bar{X} .
- Obtener la función de probabilidad de S^2 y de la cuasivarianza \hat{S}^2 .
- Comprobar que $E\bar{X} = EX$, $Var\bar{X} = \frac{VarX}{n}$, $ES^2 = \frac{n-1}{n}VarX$ y que $E(\hat{S}^2) = VarX$.

a)

| | | | | | | | | | |
|---------------|----------------|----------------|----------------|----------------|---------------|---------------|----------------|---------------|---------------|
| (X_1, X_2) | (0, 0) | (0, 1) | (0, 2) | (1, 0) | (1, 1) | (1, 2) | (2, 0) | (2, 1) | (2, 2) |
| $f(x_1, x_2)$ | $\frac{1}{36}$ | $\frac{1}{12}$ | $\frac{1}{18}$ | $\frac{1}{12}$ | $\frac{1}{4}$ | $\frac{1}{6}$ | $\frac{1}{18}$ | $\frac{1}{6}$ | $\frac{1}{9}$ |

b)

| | | | | | |
|--------------|----------------|---------------|-----------------|---------------|---------------|
| \bar{X} | 0 | $\frac{1}{2}$ | 1 | $\frac{3}{2}$ | 2 |
| $f(\bar{x})$ | $\frac{1}{36}$ | $\frac{1}{6}$ | $\frac{13}{36}$ | $\frac{1}{3}$ | $\frac{1}{9}$ |

c)

| | | | |
|----------|----------------|---------------|---------------|
| S^2 | 0 | $\frac{1}{4}$ | 1 |
| $f(s^2)$ | $\frac{7}{18}$ | $\frac{1}{2}$ | $\frac{1}{9}$ |

| | | | |
|----------------|----------------|---------------|---------------|
| \hat{S}^2 | 0 | $\frac{1}{2}$ | 2 |
| $f(\hat{s}^2)$ | $\frac{7}{18}$ | $\frac{1}{2}$ | $\frac{1}{9}$ |

d)

$$\begin{aligned}
 EX &= 0\left(\frac{1}{6}\right) + 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{3}\right) = \frac{7}{6} \\
 E\bar{X} &= 0\left(\frac{1}{36}\right) + \frac{1}{2}\left(\frac{1}{6}\right) + 1\left(\frac{13}{36}\right) + \frac{3}{2}\left(\frac{1}{3}\right) + 2\left(\frac{1}{9}\right) = \frac{7}{6} \\
 EX^2 &= 0^2\left(\frac{1}{6}\right) + 1^2\left(\frac{1}{2}\right) + 2^2\left(\frac{1}{3}\right) = \frac{11}{6} \\
 E\bar{X}^2 &= 0^2\left(\frac{1}{36}\right) + \left(\frac{1}{2}\right)^2\left(\frac{1}{6}\right) + 1^2\left(\frac{13}{36}\right) + \left(\frac{3}{2}\right)^2\left(\frac{1}{3}\right) + 2^2\left(\frac{1}{9}\right) = \frac{115}{72} \\
 VarX &= EX^2 - E^2X = \frac{17}{36} \\
 Var\bar{X} &= E\bar{X}^2 - E^2\bar{X} = \frac{17}{72} = \frac{VarX}{2} \\
 ES^2 &= 0\left(\frac{7}{18}\right) + \frac{1}{4}\left(\frac{1}{2}\right) + 1\left(\frac{1}{9}\right) = \frac{17}{72} = \frac{1}{2}VarX \\
 E\hat{S}^2 &= 0\left(\frac{7}{18}\right) + \frac{1}{2}\left(\frac{1}{2}\right) + 2\left(\frac{1}{9}\right) = \frac{17}{36} = VarX
 \end{aligned}$$

6.4 Reproductividad de distribuciones

Definición 6.2: Diremos que una cierta distribución es reproductiva respecto de un parámetro cuando al sumar variables aleatorias independientes que sigan esa misma distribución, la variable suma también va a seguir esa distribución y el parámetro actúa aditivamente. Sabemos que si X_1, X_2, \dots, X_n son variables independientes entonces:

$$\varphi_{X_1+X_2+\dots+X_n}(t) = E(e^{it(X_1+X_2+\dots+X_n)}) = \prod_{j=1}^n E(e^{itx_j}) = \prod_{j=1}^n \varphi_{X_j}(t).$$

Proposición 6.2: La distribución binomial es reproductiva respecto al primer parámetro, es decir, si X_1, X_2, \dots, X_n son variables aleatorias independientes tales que $X_j \sim \text{Bin}(m_j, p)$, $j = 1, 2, \dots, n$, entonces

$$\sum_{j=1}^n X_j \sim \text{Bin}\left(\sum_{j=1}^n m_j, p\right).$$

Demostración:

$$\varphi_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n \varphi_{X_j}(t) = \prod_{j=1}^n (pe^{it} + q)^{m_j} = (pe^{it} + q)^{\sum_{j=1}^n m_j} \sim \text{Bin}\left(\sum_{j=1}^n m_j, p\right).$$

Proposición 6.3: La distribución de Poisson es reproductiva respecto de λ , es decir, si X_1, X_2, \dots, X_n son variables aleatorias independientes tales que $X_j \sim P(\lambda_j)$, $j = 1, 2, \dots, n$, entonces

$$\sum_{j=1}^n X_j \sim P\left(\sum_{j=1}^n \lambda_j\right).$$

Demostración:

$$\varphi_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n \varphi_{X_j}(t) = \prod_{j=1}^n e^{\lambda_j(e^{it}-1)} = e^{\sum_{j=1}^n \lambda_j(e^{it}-1)} \sim P\left(\sum_{j=1}^n \lambda_j\right)$$

Proposición 6.4: La distribución Gamma es reproductiva respecto al primer parámetro, es decir, si X_1, X_2, \dots, X_n son variables aleatorias independientes tales que $X_j \sim \Gamma(a_j, p)$, $j = 1, 2, \dots, n$, entonces

$$\sum_{j=1}^n X_j \sim \Gamma\left(\sum_{j=1}^n a_j, p\right).$$

Demostración:

$$\varphi_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n \varphi_{X_j}(t) = \prod_{j=1}^n \left(\frac{p}{p-it}\right)^{a_j} = \left(\frac{p}{p-it}\right)^{\sum_{j=1}^n a_j} \sim \Gamma\left(\sum_{j=1}^n a_j, p\right)$$

Proposición 6.5: Sean X_1, X_2, \dots, X_n variables independientes tales que $X_j \sim N(\mu_j, \sigma_j)$, $j = 1, 2, \dots, n$, y sean a su vez a_1, a_2, \dots, a_n y $b \in \mathbb{R}$ entonces

$$\sum_{j=1}^n a_j X_j + b \sim N\left(\sum_{j=1}^n a_j \mu_j + b, \sqrt{\sum_{j=1}^n a_j^2 \sigma_j^2}\right)$$

Demostración:

$$\begin{aligned} \varphi_{\sum_{j=1}^n a_j X_j + b}(t) &= E[e^{it(\sum_{j=1}^n a_j X_j + b)}] = e^{itb} \prod_{j=1}^n E(e^{ita_j X_j}) \\ &= e^{itb} \prod_{j=1}^n \varphi_{X_j}(ta_j) = e^{itb} \prod_{j=1}^n [e^{ita_j \mu_j - \frac{1}{2} \sigma_j^2 a_j^2 t^2}] \end{aligned}$$

$$= e^{itb} \left(\sum_j \mu_j + b - \frac{1}{2} \sigma_j^2 a_j^2 t^2 \right) \sim N \left(\sum_{j=1}^n a_j \mu_j + b, \sqrt{\sum_{j=1}^n a_j^2 \sigma_j^2} \right).$$

Ejemplo 6.7: Calcular la distribución exacta del estadístico $\sum_{i=1}^n X_i$ obtenido a partir de muestras de tamaño n de una población X tal que:

- a) $X \rightsquigarrow N(0, 1)$.
- b) $X \rightsquigarrow \exp(\lambda)$.
- c) $X \rightsquigarrow B(m, p)$.
- d) $X \rightsquigarrow P(\lambda)$.
- e) $X \rightsquigarrow \Gamma(a, p)$.

Sea $T = \sum_{i=1}^n X_i$

a) $T \rightsquigarrow N(ET, \sqrt{\text{Var}(T)})$, donde

$$ET = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = n(0) = 0,$$

$$\text{Var}(T) = \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) = n(1) = n.$$

Por lo que finalmente $T \rightsquigarrow N(0, \sqrt{n})$.

b) Sabemos que la distribución Gamma es reproductiva respecto del primer parámetro. Por tanto $T \rightsquigarrow \Gamma(n, \lambda)$. Por ser la suma de $n \exp(\lambda) \equiv \Gamma(1, \lambda)$.

c) Sabemos por la reproductividad de la distribución Binomial que si $T = \sum_{i=1}^n X_i$, entonces, $T \rightsquigarrow \text{Bin}(nm, p)$

d) Sabemos por la reproductividad de la distribución de Poisson que si $T = \sum_{i=1}^n X_i$, entonces, $T \rightsquigarrow P(n\lambda)$.

e) Sabemos por la reproductividad respecto del primer parámetro de la función de Gamma que si $T = \sum_{i=1}^n X_i$, entonces, $T \rightsquigarrow \Gamma(na, p)$.

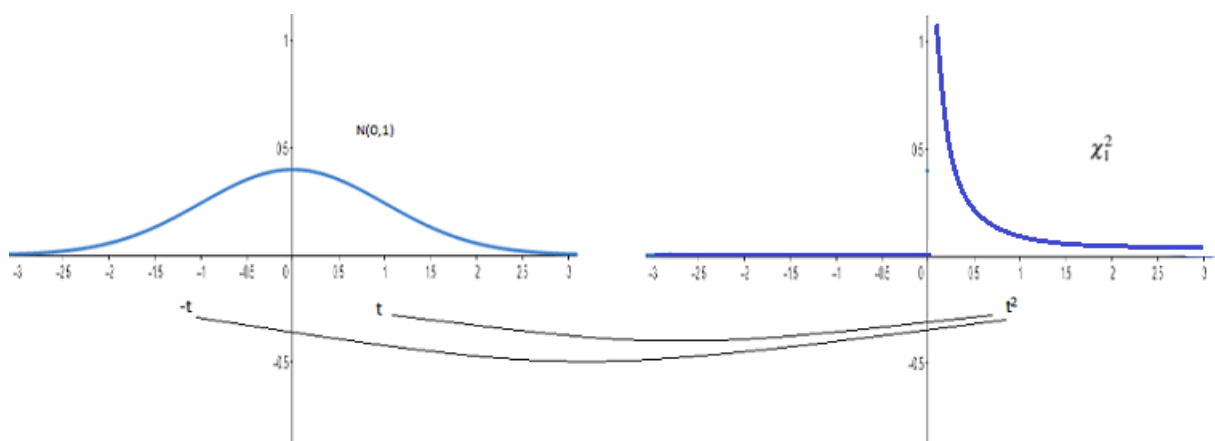
6.5 Distribuciones asociadas al muestreo

En este apartado introducimos tres distribuciones nuevas de amplia utilidad en la inferencia estadística:

- i) Distribución ji-cuadrado (o chi-cuadrado).
- ii) Distribución t de Student.
- iii) Distribución F de Snedecor.

Distribución Ji-cuadrado

Definición 6.3: Sea $X \sim N(0, 1)$ y sea $Y = X^2$. Por definición diremos que Y sigue una distribución ji-cuadrado con un grado de libertad. Lo notaremos $Y \sim \chi_1^2$.



La siguiente proposición relaciona la distribución ji-cuadrado con la distribución Gamma.

Proposición 6.6: Si $Y \sim \chi_1^2$, entonces $Y \sim \Gamma(\frac{1}{2}, \frac{1}{2})$.

Demostración: Si $Y \sim \chi_1^2$, entonces $Y = X^2$ con $X \sim N(0, 1)$. La transformación $Y = X^2$ no es biyectiva. Entonces para determinar la densidad de Y no podemos aplicar el teorema Jacobiano. Calculamos la función de distribución de Y . Si $y > 0$

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = 2p(0 \leq X \leq \sqrt{y}) = 2 \int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

Haciendo uso del teorema fundamental del cálculo¹

$$f_Y(y) = F'_Y(y) = 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \frac{1}{2\sqrt{y}} \text{ si } y > 0,$$

¹Teorema fundamental del cálculo:

$$G(y) = \int_a^{g(y)} f(x) dx \implies G'(y) = f(g(y))g'(y).$$

luego:

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} y^{-\frac{1}{2}} & \text{si } y > 0 \\ 0 & \text{fuera.} \end{cases}$$

Además

$$\begin{aligned} f_{\Gamma(\frac{1}{2}, \frac{1}{2})}(y) &= \begin{cases} \frac{(\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} e^{-\frac{y}{2}} y^{-\frac{1}{2}} & \text{si } y > 0 \\ 0 & \text{fuera} \end{cases} \\ &= f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} y^{-\frac{1}{2}} & \text{si } y > 0 \\ 0 & \text{fuera.} \end{cases} \end{aligned}$$

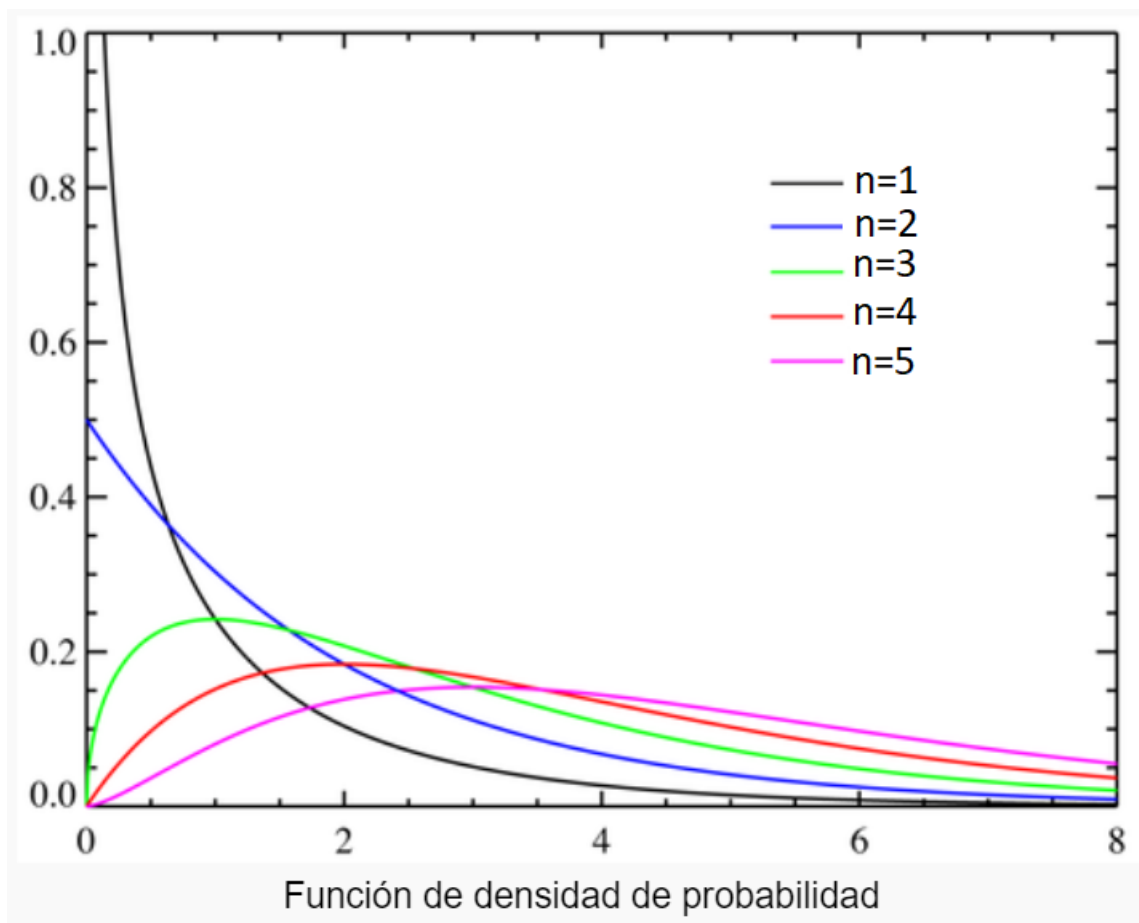
Definición 6.4: Sea X_1, \dots, X_n una muestra aleatoria simple de $N(0, 1)$ y sea $Y = \sum_{i=1}^n X_i^2$. Por definición diremos que Y sigue una distribución ji-cuadrado con n grados de libertad. Lo notaremos $Y \sim \chi_n^2$

Proposición 6.7: Si $Y \sim \chi_n^2$, entonces $Y \sim \Gamma(\frac{n}{2}, \frac{1}{2})$.

Demostración: Si $Y \sim \chi_n^2$ entonces $Y = \sum_{i=1}^n X_i^2$ con X_1, \dots, X_n variables aleatorias independientes igualmente distribuidas según una $N(0, 1)$. Además cada X_i^2 es una $\Gamma(\frac{1}{2}, \frac{1}{2})$. Así, Y es la suma de n distribuciones $\Gamma(\frac{1}{2}, \frac{1}{2})$ independientes. Como la distribución Gamma es reproductiva respecto al primer parámetro, $Y \sim \Gamma(\frac{n}{2}, \frac{1}{2})$.

La función de densidad de una ji-cuadrado con n grados de libertad es:

$$f_{\chi_n^2}(y) = \begin{cases} \frac{(\frac{1}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} e^{-\frac{y}{2}} y^{\frac{n}{2}-1} & \text{si } y > 0 \\ 0 & \text{fuera.} \end{cases}$$



La esperanza de una ji-cuadrado con n grados de libertad es:

$$E(\chi_n^2) = E\left[\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)\right] = \frac{\frac{n}{2}}{\frac{1}{2}} = n,$$

y la varianza:

$$\text{Var}(\chi_n^2) = \text{Var}\left[\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)\right] = \frac{\frac{n}{2}}{\left(\frac{1}{2}\right)^2} = 2n$$

La curva de densidad, en general, no admite primitiva y los valores de la distribución están tabulados.

Distribución Chi Cuadrado χ^2

P = Probabilidad de encontrar un valor mayor o igual que el chi cuadrado tabulado, v = Grados de Libertad

| v/p | 0.001 | 0.0025 | 0.005 | 0.01 | 0.025 | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.35 | 0.4 | 0.45 | 0.5 |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1 | 10.8274 | 9.1404 | 7.8794 | 6.6349 | 5.0239 | 3.8415 | 2.7055 | 2.0722 | 1.6424 | 1.3233 | 1.0742 | 0.8735 | 0.7083 | 0.5707 | 0.4549 |
| 2 | 13.8150 | 11.9827 | 10.5965 | 9.2104 | 7.3778 | 5.9915 | 4.6052 | 3.7942 | 3.2189 | 2.7726 | 2.4079 | 2.0996 | 1.8326 | 1.5970 | 1.3863 |
| 3 | 16.2660 | 14.3202 | 12.8381 | 11.3449 | 9.3484 | 7.8147 | 6.2514 | 5.3170 | 4.6416 | 4.1083 | 3.6649 | 3.2831 | 2.9462 | 2.6430 | 2.3660 |
| 4 | 18.4662 | 16.4238 | 14.8602 | 13.2767 | 11.1433 | 9.4877 | 7.7794 | 6.7449 | 5.9886 | 5.3853 | 4.8784 | 4.4377 | 4.0446 | 3.6871 | 3.3567 |
| 5 | 20.5147 | 18.3854 | 16.7496 | 15.0863 | 12.8325 | 11.0705 | 9.2363 | 8.1152 | 7.2893 | 6.6257 | 6.0644 | 5.5731 | 5.1319 | 4.7278 | 4.3515 |
| 6 | 22.4575 | 20.2491 | 18.5475 | 16.8119 | 14.4494 | 12.5916 | 10.6446 | 9.4461 | 8.5581 | 7.8408 | 7.2311 | 6.6948 | 6.2108 | 5.7652 | 5.3481 |
| 7 | 24.3213 | 22.0402 | 20.2777 | 18.4753 | 16.0128 | 14.0671 | 12.0170 | 10.7479 | 9.8032 | 9.0371 | 8.3834 | 7.8061 | 7.2832 | 6.8000 | 6.3458 |
| 8 | 26.1239 | 23.7742 | 21.9549 | 20.0902 | 17.5345 | 15.5073 | 13.3616 | 12.0271 | 11.0301 | 10.2189 | 9.5245 | 8.9094 | 8.3505 | 7.8325 | 7.3441 |
| 9 | 27.8767 | 25.4625 | 23.5893 | 21.6660 | 19.0228 | 16.9190 | 14.6837 | 13.2880 | 12.2421 | 11.3887 | 10.6564 | 10.0060 | 9.4136 | 8.8632 | 8.3428 |
| 10 | 29.5879 | 27.1119 | 25.1881 | 23.2093 | 20.4832 | 18.3070 | 15.9872 | 14.5339 | 13.4420 | 12.5489 | 11.7807 | 11.0971 | 10.4732 | 9.8922 | 9.3418 |
| 11 | 31.2635 | 28.7291 | 26.7569 | 24.7250 | 21.9200 | 19.6752 | 17.2750 | 15.7671 | 14.6314 | 13.7007 | 12.8987 | 12.1836 | 11.5298 | 10.9199 | 10.3410 |
| 12 | 32.9092 | 30.3182 | 28.2997 | 26.2170 | 23.3367 | 21.0261 | 18.5493 | 16.9893 | 15.8120 | 14.8454 | 14.0111 | 13.2661 | 12.5838 | 11.9463 | 11.3403 |
| 13 | 34.5274 | 31.8830 | 29.8193 | 27.6882 | 24.7356 | 22.3620 | 19.8119 | 18.2020 | 16.9848 | 15.9839 | 15.1187 | 14.3451 | 13.6356 | 12.9717 | 12.3398 |
| 14 | 36.1239 | 33.4262 | 31.3194 | 29.1412 | 26.1189 | 23.6848 | 21.0641 | 19.4062 | 18.1508 | 17.1169 | 16.2221 | 15.4209 | 14.6853 | 13.9961 | 13.3393 |
| 15 | 37.6978 | 34.9494 | 32.8015 | 30.5780 | 27.4884 | 24.9958 | 22.3071 | 20.6030 | 19.3107 | 18.2451 | 17.3217 | 16.4940 | 15.7332 | 15.0197 | 14.3389 |
| 16 | 39.2518 | 36.4555 | 34.2671 | 31.9999 | 28.8453 | 26.2962 | 23.5418 | 21.7931 | 20.4651 | 19.3689 | 18.4179 | 17.5646 | 16.7795 | 16.0425 | 15.3385 |
| 17 | 40.7911 | 37.9462 | 35.7184 | 33.4087 | 30.1910 | 27.5871 | 24.7690 | 22.9770 | 21.6146 | 20.4887 | 19.5110 | 18.6330 | 17.8244 | 17.0646 | 16.3382 |
| 18 | 42.3119 | 39.4220 | 37.1564 | 34.8052 | 31.5264 | 28.8693 | 25.9894 | 24.1555 | 22.7595 | 21.6049 | 20.6014 | 19.6993 | 18.8679 | 18.0860 | 17.3379 |
| 19 | 43.8194 | 40.8847 | 38.5821 | 36.1908 | 32.8523 | 30.1435 | 27.2036 | 25.3289 | 23.9004 | 22.7178 | 21.6891 | 20.7638 | 19.9102 | 19.1069 | 18.3376 |
| 20 | 45.3142 | 42.3358 | 39.9969 | 37.5663 | 34.1696 | 31.4104 | 28.4120 | 26.4976 | 25.0375 | 23.8277 | 22.7745 | 21.8265 | 20.9514 | 20.1272 | 19.3374 |
| 21 | 46.7963 | 43.7749 | 41.4009 | 38.9322 | 35.4789 | 32.6706 | 29.6151 | 27.6620 | 26.1711 | 24.9348 | 23.8578 | 22.8876 | 21.9915 | 21.1470 | 20.3372 |
| 22 | 48.2676 | 45.2041 | 42.7957 | 40.2894 | 36.7807 | 33.9245 | 30.8133 | 28.8224 | 27.3015 | 26.0393 | 24.9390 | 23.9473 | 23.0307 | 22.1663 | 21.3370 |
| 23 | 49.7276 | 46.6231 | 44.1814 | 41.6383 | 38.0756 | 35.1725 | 32.0069 | 29.9792 | 28.4288 | 27.1413 | 26.0184 | 25.0055 | 24.0689 | 23.1852 | 22.3369 |
| 24 | 51.1790 | 48.0336 | 45.5584 | 42.9798 | 39.3641 | 36.4150 | 33.1962 | 31.1325 | 29.5533 | 28.2412 | 27.0960 | 26.0625 | 25.1064 | 24.2037 | 23.3367 |
| 25 | 52.6187 | 49.4351 | 46.9280 | 44.3140 | 40.6465 | 37.6525 | 34.3816 | 32.2825 | 30.6752 | 29.3388 | 28.1719 | 27.1183 | 26.1430 | 25.2218 | 24.3366 |
| 26 | 54.0511 | 50.8291 | 48.2898 | 45.6416 | 41.9231 | 38.8851 | 35.5632 | 33.4295 | 31.7946 | 30.4346 | 29.2463 | 28.1730 | 27.1789 | 26.2395 | 25.3365 |
| 27 | 55.4751 | 52.2152 | 49.6450 | 46.9628 | 43.1945 | 40.1133 | 36.7412 | 34.5736 | 32.9117 | 31.5284 | 30.3193 | 29.2266 | 28.2141 | 27.2569 | 26.3363 |
| 28 | 56.8918 | 53.5939 | 50.9936 | 48.2782 | 44.4608 | 41.3372 | 37.9159 | 35.7150 | 34.0266 | 32.6205 | 31.3909 | 30.2791 | 29.2486 | 28.2740 | 27.3362 |
| 29 | 58.3006 | 54.9662 | 52.3355 | 49.5878 | 45.7223 | 42.5569 | 39.0875 | 36.8538 | 35.1394 | 33.7109 | 32.4612 | 31.3308 | 30.2825 | 29.2908 | 28.3361 |

Distribución T de Student.

Definición 6.5: Sea $X \sim N(0, 1)$ e $Y \sim \chi_n^2$ ambas independientes y sea $T = \frac{X}{\sqrt{\frac{Y}{n}}}$. Por definición T sigue una distribución t de Student con n grados de libertad. Lo notaremos como $T \sim t_n$.

Por tanto, siempre que dividamos una normal tipificada entre la raíz de una χ^2 corregida por sus grados de libertad, siendo ambas independientes, la distribución resultante es una t .

La forma de la t es similar a la de una $N(0, 1)$ aunque con las colas más pesadas. Su función de densidad es

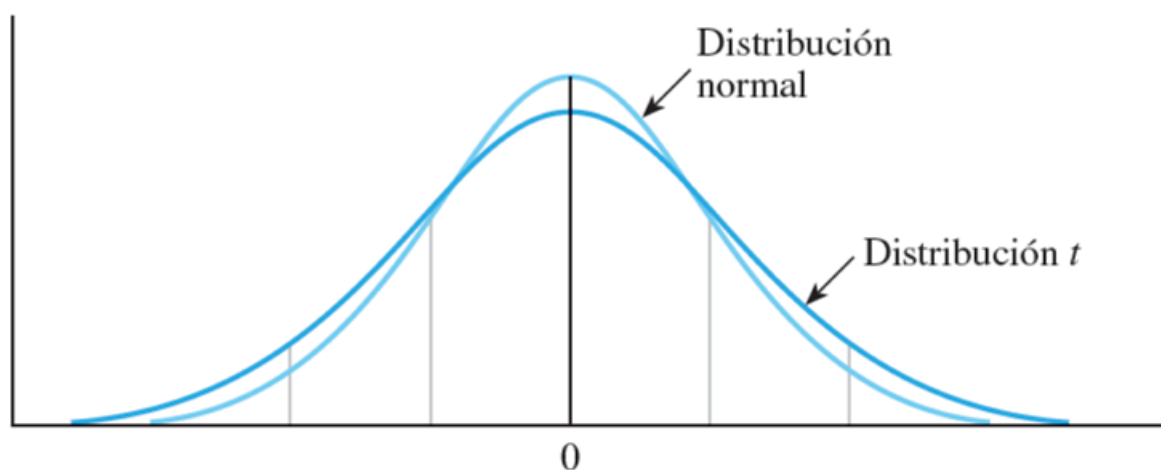
$$f_{t_n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}},$$

para $t \in \mathbb{R}$, donde n denota los grados de libertad y Γ es la función gamma.

La expresión anterior también suele escribirse como

$$f_{t_n}(t) = \frac{1}{\sqrt{n} B\left(\frac{1}{2}, \frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}},$$

para $t \in \mathbb{R}$, donde B es la función beta.



Por tanto, $E(t_n) = 0$ para todo $n > 1$ y la varianza de t_n para valores $n > 2$ es $\text{Var}(X) = \frac{n}{n-2} \cdot t_1$ es la distribución de Cauchy² y no existe $E(t_1)$.

No existe, en general, primitiva de $f_{t_n}(t)$ y, por tanto, las probabilidades están tabuladas.

²Sea X una variable aleatoria con distribución de Cauchy, su densidad es

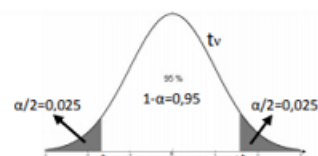
$$f_X(x) = \begin{cases} \frac{1}{\pi(1+x^2)} & \text{si } x \in \mathbb{R} \\ 0 & \text{fuera.} \end{cases}$$

La distribución de Cauchy tiene las colas más pesadas que la de la $N(0, 1)$ hasta el punto de que no tiene EX .

$$EX = \int_{-\infty}^{\infty} x \left(\frac{1}{\pi(1+x^2)} \right) dx = \frac{1}{2\pi} [\ln(1+x^2)]_{-\infty}^{\infty} = \frac{1}{2\pi} (\ln(\infty) - \ln(\infty)).$$

Distribución t de Student

Contiene los valores de t tales que $\frac{\alpha}{2} = P(t_v \geq t)$, donde v son los Grados de Libertad



| | | $\alpha/2$ | | | | | | | | | | | | |
|-------------------------|----|------------|---------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|-------|
| | | 0,0005 | 0,001 | 0,005 | 0,01 | 0,025 | 0,05 | 0,1 | 0,2 | 0,25 | 0,3 | 0,4 | 0,45 | 0,475 |
| v grados de libertad | 1 | 636,619 | 318,309 | 63,657 | 31,821 | 12,706 | 6,314 | 3,078 | 1,376 | 1,000 | 0,727 | 0,325 | 0,158 | 0,079 |
| | 2 | 31,599 | 22,327 | 9,925 | 6,965 | 4,303 | 2,920 | 1,886 | 1,061 | 0,816 | 0,617 | 0,289 | 0,142 | 0,071 |
| | 3 | 12,924 | 10,215 | 5,841 | 4,541 | 3,182 | 2,353 | 1,638 | 0,978 | 0,765 | 0,584 | 0,277 | 0,137 | 0,068 |
| | 4 | 8,610 | 7,173 | 4,604 | 3,747 | 2,776 | 2,132 | 1,533 | 0,941 | 0,741 | 0,569 | 0,271 | 0,134 | 0,067 |
| | 5 | 6,869 | 5,893 | 4,032 | 3,365 | 2,571 | 2,015 | 1,476 | 0,920 | 0,727 | 0,559 | 0,267 | 0,132 | 0,066 |
| | 6 | 5,959 | 5,208 | 3,707 | 3,143 | 2,447 | 1,943 | 1,440 | 0,906 | 0,718 | 0,553 | 0,265 | 0,131 | 0,065 |
| | 7 | 5,408 | 4,785 | 3,499 | 2,998 | 2,365 | 1,895 | 1,415 | 0,896 | 0,711 | 0,549 | 0,263 | 0,130 | 0,065 |
| | 8 | 5,041 | 4,501 | 3,355 | 2,896 | 2,306 | 1,860 | 1,397 | 0,889 | 0,706 | 0,546 | 0,262 | 0,130 | 0,065 |
| | 9 | 4,781 | 4,297 | 3,250 | 2,821 | 2,262 | 1,833 | 1,383 | 0,883 | 0,703 | 0,543 | 0,261 | 0,129 | 0,064 |
| | 10 | 4,587 | 4,144 | 3,169 | 2,764 | 2,228 | 1,812 | 1,372 | 0,879 | 0,700 | 0,542 | 0,260 | 0,129 | 0,064 |
| | 11 | 4,437 | 4,025 | 3,106 | 2,718 | 2,201 | 1,796 | 1,363 | 0,876 | 0,697 | 0,540 | 0,260 | 0,129 | 0,064 |
| | 12 | 4,318 | 3,930 | 3,055 | 2,681 | 2,179 | 1,782 | 1,356 | 0,873 | 0,695 | 0,539 | 0,259 | 0,128 | 0,064 |
| | 13 | 4,221 | 3,852 | 3,012 | 2,650 | 2,160 | 1,771 | 1,350 | 0,870 | 0,694 | 0,538 | 0,259 | 0,128 | 0,064 |
| | 14 | 4,140 | 3,787 | 2,977 | 2,624 | 2,145 | 1,761 | 1,345 | 0,868 | 0,692 | 0,537 | 0,258 | 0,128 | 0,064 |
| | 15 | 4,073 | 3,733 | 2,947 | 2,602 | 2,131 | 1,753 | 1,341 | 0,866 | 0,691 | 0,536 | 0,258 | 0,128 | 0,064 |
| | 16 | 4,015 | 3,686 | 2,921 | 2,583 | 2,120 | 1,746 | 1,337 | 0,865 | 0,690 | 0,535 | 0,258 | 0,128 | 0,064 |
| | 17 | 3,965 | 3,646 | 2,898 | 2,567 | 2,110 | 1,740 | 1,333 | 0,863 | 0,689 | 0,534 | 0,257 | 0,128 | 0,064 |
| | 18 | 3,922 | 3,610 | 2,878 | 2,552 | 2,101 | 1,734 | 1,330 | 0,862 | 0,688 | 0,534 | 0,257 | 0,127 | 0,064 |
| | 19 | 3,883 | 3,579 | 2,861 | 2,539 | 2,093 | 1,729 | 1,328 | 0,861 | 0,688 | 0,533 | 0,257 | 0,127 | 0,064 |
| | 20 | 3,850 | 3,552 | 2,845 | 2,528 | 2,086 | 1,725 | 1,325 | 0,860 | 0,687 | 0,533 | 0,257 | 0,127 | 0,063 |
| | 21 | 3,819 | 3,527 | 2,831 | 2,518 | 2,080 | 1,721 | 1,323 | 0,859 | 0,686 | 0,532 | 0,257 | 0,127 | 0,063 |
| | 22 | 3,792 | 3,505 | 2,819 | 2,508 | 2,074 | 1,717 | 1,321 | 0,858 | 0,686 | 0,532 | 0,256 | 0,127 | 0,063 |
| | 23 | 3,768 | 3,485 | 2,807 | 2,500 | 2,069 | 1,714 | 1,319 | 0,858 | 0,685 | 0,532 | 0,256 | 0,127 | 0,063 |
| | 24 | 3,745 | 3,467 | 2,797 | 2,492 | 2,064 | 1,711 | 1,318 | 0,857 | 0,685 | 0,531 | 0,256 | 0,127 | 0,063 |
| | 25 | 3,725 | 3,450 | 2,787 | 2,485 | 2,060 | 1,708 | 1,316 | 0,856 | 0,684 | 0,531 | 0,256 | 0,127 | 0,063 |
| | 26 | 3,707 | 3,435 | 2,779 | 2,479 | 2,056 | 1,706 | 1,315 | 0,856 | 0,684 | 0,531 | 0,256 | 0,127 | 0,063 |
| | 27 | 3,690 | 3,421 | 2,771 | 2,473 | 2,052 | 1,703 | 1,314 | 0,855 | 0,684 | 0,531 | 0,256 | 0,127 | 0,063 |
| | 28 | 3,674 | 3,408 | 2,763 | 2,467 | 2,048 | 1,701 | 1,313 | 0,855 | 0,683 | 0,530 | 0,256 | 0,127 | 0,063 |
| | 29 | 3,659 | 3,396 | 2,756 | 2,462 | 2,045 | 1,699 | 1,311 | 0,854 | 0,683 | 0,530 | 0,256 | 0,127 | 0,063 |
| | 30 | 3,646 | 3,385 | 2,750 | 2,457 | 2,042 | 1,697 | 1,310 | 0,854 | 0,683 | 0,530 | 0,256 | 0,127 | 0,063 |
| | 31 | 3,633 | 3,375 | 2,744 | 2,453 | 2,040 | 1,696 | 1,309 | 0,853 | 0,682 | 0,530 | 0,256 | 0,127 | 0,063 |
| | 32 | 3,622 | 3,365 | 2,738 | 2,449 | 2,037 | 1,694 | 1,309 | 0,853 | 0,682 | 0,530 | 0,255 | 0,127 | 0,063 |
| | 33 | 3,611 | 3,356 | 2,733 | 2,445 | 2,035 | 1,692 | 1,308 | 0,853 | 0,682 | 0,530 | 0,255 | 0,127 | 0,063 |
| | 34 | 3,601 | 3,348 | 2,728 | 2,441 | 2,032 | 1,691 | 1,307 | 0,852 | 0,682 | 0,529 | 0,255 | 0,127 | 0,063 |
| | 35 | 3,591 | 3,340 | 2,724 | 2,438 | 2,030 | 1,690 | 1,306 | 0,852 | 0,682 | 0,529 | 0,255 | 0,127 | 0,063 |
| | | α | 0,001 | 0,002 | 0,01 | 0,02 | 0,05 | 0,1 | 0,2 | 0,4 | 0,5 | 0,6 | 0,8 | 0,9 |

| | | $\alpha/2$ | | | | | | | | | | | | |
|-------------------------|----------|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | | 0,0005 | 0,001 | 0,005 | 0,01 | 0,025 | 0,05 | 0,1 | 0,2 | 0,25 | 0,3 | 0,4 | 0,45 | 0,475 |
| v grados de libertad | 36 | 3,582 | 3,333 | 2,719 | 2,434 | 2,028 | 1,688 | 1,306 | 0,852 | 0,681 | 0,529 | 0,255 | 0,127 | 0,063 |
| | 37 | 3,574 | 3,326 | 2,715 | 2,431 | 2,026 | 1,687 | 1,305 | 0,851 | 0,681 | 0,529 | 0,255 | 0,127 | 0,063 |
| | 38 | 3,566 | 3,319 | 2,712 | 2,429 | 2,024 | 1,686 | 1,304 | 0,851 | 0,681 | 0,529 | 0,255 | 0,127 | 0,063 |
| | 39 | 3,558 | 3,313 | 2,708 | 2,426 | 2,023 | 1,685 | 1,304 | 0,851 | 0,681 | 0,529 | 0,255 | 0,126 | 0,063 |
| | 40 | 3,551 | 3,307 | 2,704 | 2,423 | 2,021 | 1,684 | 1,303 | 0,851 | 0,681 | 0,529 | 0,255 | 0,126 | 0,063 |
| | 41 | 3,544 | 3,301 | 2,701 | 2,421 | 2,020 | 1,683 | 1,303 | 0,850 | 0,681 | 0,529 | 0,255 | 0,126 | 0,063 |
| | 42 | 3,538 | 3,296 | 2,698 | 2,418 | 2,018 | 1,682 | 1,302 | 0,850 | 0,680 | 0,528 | 0,255 | 0,126 | 0,063 |
| | 43 | 3,532 | 3,291 | 2,695 | 2,416 | 2,017 | 1,681 | 1,302 | 0,850 | 0,680 | 0,528 | 0,255 | 0,126 | 0,063 |
| | 44 | 3,526 | 3,286 | 2,692 | 2,414 | 2,015 | 1,680 | 1,301 | 0,850 | 0,680 | 0,528 | 0,255 | 0,126 | 0,063 |
| | 45 | 3,520 | 3,281 | 2,690 | 2,412 | 2,014 | 1,679 | 1,301 | 0,850 | 0,680 | 0,528 | 0,255 | 0,126 | 0,063 |
| | 46 | 3,515 | 3,277 | 2,687 | 2,410 | 2,013 | 1,679 | 1,300 | 0,850 | 0,680 | 0,528 | 0,255 | 0,126 | 0,063 |
| | 47 | 3,510 | 3,273 | 2,685 | 2,408 | 2,012 | 1,678 | 1,300 | 0,849 | 0,680 | 0,528 | 0,255 | 0,126 | 0,063 |
| | 48 | 3,505 | 3,269 | 2,682 | 2,407 | 2,011 | 1,677 | 1,299 | 0,849 | 0,680 | 0,528 | 0,255 | 0,126 | 0,063 |
| | 49 | 3,500 | 3,265 | 2,680 | 2,405 | 2,010 | 1,677 | 1,299 | 0,849 | 0,680 | 0,528 | 0,255 | 0,126 | 0,063 |
| | 50 | 3,496 | 3,261 | 2,678 | 2,403 | 2,009 | 1,676 | 1,299 | 0,849 | 0,679 | 0,528 | 0,255 | 0,126 | 0,063 |
| | 51 | 3,492 | 3,258 | 2,676 | 2,402 | 2,008 | 1,675 | 1,298 | 0,849 | 0,679 | 0,528 | 0,255 | 0,126 | 0,063 |
| | 52 | 3,488 | 3,255 | 2,674 | 2,400 | 2,007 | 1,675 | 1,298 | 0,849 | 0,679 | 0,528 | 0,255 | 0,126 | 0,063 |
| | 53 | 3,484 | 3,251 | 2,672 | 2,399 | 2,006 | 1,674 | 1,298 | 0,848 | 0,679 | 0,528 | 0,255 | 0,126 | 0,063 |
| | 54 | 3,480 | 3,248 | 2,670 | 2,397 | 2,005 | 1,674 | 1,297 | 0,848 | 0,679 | 0,528 | 0,255 | 0,126 | 0,063 |
| | 55 | 3,476 | 3,245 | 2,668 | 2,396 | 2,004 | 1,673 | 1,297 | 0,848 | 0,679 | 0,527 | 0,255 | 0,126 | 0,063 |
| | 56 | 3,473 | 3,242 | 2,667 | 2,395 | 2,003 | 1,673 | 1,297 | 0,848 | 0,679 | 0,527 | 0,255 | 0,126 | 0,063 |
| | 57 | 3,470 | 3,239 | 2,665 | 2,394 | 2,002 | 1,672 | 1,297 | 0,848 | 0,679 | 0,527 | 0,255 | 0,126 | 0,063 |
| | 58 | 3,466 | 3,237 | 2,663 | 2,392 | 2,002 | 1,672 | 1,296 | 0,848 | 0,679 | 0,527 | 0,255 | 0,126 | 0,063 |
| | 59 | 3,463 | 3,234 | 2,662 | 2,391 | 2,001 | 1,671 | 1,296 | 0,848 | 0,679 | 0,527 | 0,254 | 0,126 | 0,063 |
| | 60 | 3,460 | 3,232 | 2,660 | 2,390 | 2,000 | 1,671 | 1,296 | 0,848 | 0,679 | 0,527 | 0,254 | 0,126 | 0,063 |
| | 120 | 3,373 | 3,160 | 2,617 | 2,358 | 1,980 | 1,658 | 1,289 | 0,845 | 0,677 | 0,526 | 0,254 | 0,126 | 0,063 |
| | ∞ | 3,300 | 3,098 | 2,581 | 2,330 | 1,962 | 1,646 | 1,282 | 0,842 | 0,675 | 0,525 | 0,253 | 0,126 | 0,063 |
| | | α | 0.001 | 0.002 | 0.01 | 0.02 | 0.05 | 0.1 | 0.2 | 0.4 | 0.5 | 0.6 | 0.8 | 0.9 |

Distribución F de Snedecor.

Definición 6.6: Sea $X \sim \chi_m^2$ e $Y \sim \chi_n^2$ ambas independientes, definimos:

$$F = \frac{\frac{X}{m}}{\frac{Y}{n}}.$$

Por definición diremos que F sigue una distribución F de Snedecor con m y n grados de libertad (ó m grados en el numerador y n en el denominador). Lo notaremos como $F \sim F_{m,n}$.

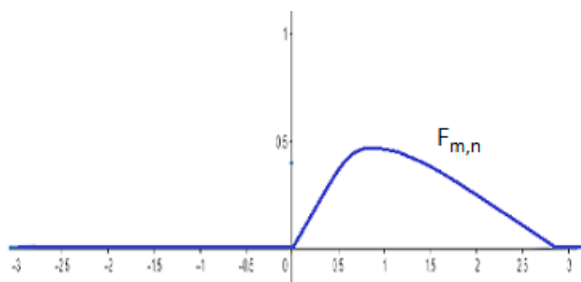
Por lo tanto, siempre que dividamos dos ji-cuadrado independientes entre sí, corregidas por sus grados de libertad obtendremos una distribución F de Snedecor. La densidad de una $F_{m,n}$ es

$$f_{F_{m,n}}(x) = \begin{cases} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{x^{\frac{m-2}{2}}}{(1+\frac{mx}{n})^{\frac{m+n}{2}}} & \text{si } x > 0 \\ 0 & \text{fuera.} \end{cases}$$

La expresión anterior también suele escribirse como

$$f_{F_{m,n}}(x) = \begin{cases} f(x) = \frac{1}{xB\left(\frac{m}{2}, \frac{n}{2}\right)} \sqrt{\frac{(mx)^m n^n}{(mx+n)^{m+n}}} & \text{si } x > 0 \\ 0 & \text{fuera,} \end{cases}$$

donde B es la función beta.



La media de $F_{m,n}$ es

$$E(X) = \frac{n}{n-2}$$

para $n > 2$ y la varianza de $F_{m,n}$ está dada por

$$\text{Var}(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$$

para $n > 4$.

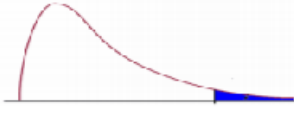
La densidad no es integrable, por lo que las probabilidades están tabuladas. Para cada probabilidad prefijada es necesaria una tabla. Esto es así porque tanto las filas como las columnas se necesitan para los grados de libertad del numerador y del denominador.

Distribución F 0.05

En las columnas se encuentran los valores F que corresponden al área 0.05 a la derecha

En las columnas se encuentran los grados de libertad del numerador

En los renglones se encuentran los grados de libertad del denominador.



| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 161.4 | 199.5 | 215.7 | 224.6 | 230.2 | 234.0 | 236.8 | 238.9 | 240.5 | 241.9 | 243.0 | 243.9 | 245.9 | 248.0 | 249.1 | 250.1 | 251.1 | 252.2 | 253.3 |
| 2 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 | 19.40 | 19.40 | 19.41 | 19.43 | 19.45 | 19.45 | 19.46 | 19.47 | 19.48 | 19.49 |
| 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 | 8.79 | 8.76 | 8.74 | 8.70 | 8.66 | 8.64 | 8.62 | 8.59 | 8.57 | 8.55 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 | 5.96 | 5.94 | 5.91 | 5.86 | 5.80 | 5.77 | 5.75 | 5.72 | 5.69 | 5.66 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 | 4.74 | 4.70 | 4.68 | 4.62 | 4.56 | 4.53 | 4.50 | 4.46 | 4.43 | 4.40 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 | 4.06 | 4.03 | 4.00 | 3.94 | 3.87 | 3.84 | 3.81 | 3.77 | 3.74 | 3.70 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 | 3.64 | 3.60 | 3.57 | 3.51 | 3.44 | 3.41 | 3.38 | 3.34 | 3.30 | 3.27 |
| 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 | 3.35 | 3.31 | 3.28 | 3.22 | 3.15 | 3.12 | 3.08 | 3.04 | 3.01 | 2.97 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 | 3.14 | 3.10 | 3.07 | 3.01 | 2.94 | 2.90 | 2.86 | 2.83 | 2.79 | 2.75 |
| 10 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 | 2.98 | 2.94 | 2.91 | 2.85 | 2.77 | 2.74 | 2.70 | 2.66 | 2.62 | 2.58 |
| 11 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.90 | 2.85 | 2.82 | 2.79 | 2.72 | 2.65 | 2.61 | 2.57 | 2.53 | 2.49 | 2.45 |
| 12 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 | 2.75 | 2.72 | 2.69 | 2.62 | 2.54 | 2.51 | 2.47 | 2.43 | 2.38 | 2.34 |
| 13 | 4.67 | 3.81 | 3.41 | 3.18 | 3.03 | 2.92 | 2.83 | 2.77 | 2.71 | 2.67 | 2.63 | 2.60 | 2.53 | 2.46 | 2.42 | 2.38 | 2.34 | 2.30 | 2.25 |
| 14 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.76 | 2.70 | 2.65 | 2.60 | 2.57 | 2.53 | 2.46 | 2.39 | 2.35 | 2.31 | 2.27 | 2.22 | 2.18 |
| 15 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 | 2.54 | 2.51 | 2.48 | 2.40 | 2.33 | 2.29 | 2.25 | 2.20 | 2.16 | 2.11 |
| 16 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.54 | 2.49 | 2.46 | 2.42 | 2.35 | 2.28 | 2.24 | 2.19 | 2.15 | 2.11 | 2.06 |
| 17 | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 | 2.55 | 2.49 | 2.45 | 2.41 | 2.38 | 2.31 | 2.23 | 2.19 | 2.15 | 2.10 | 2.06 | 2.01 |
| 18 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.58 | 2.51 | 2.46 | 2.41 | 2.37 | 2.34 | 2.27 | 2.19 | 2.15 | 2.11 | 2.06 | 2.02 | 1.97 |
| 19 | 4.38 | 3.52 | 3.13 | 2.90 | 2.74 | 2.63 | 2.54 | 2.48 | 2.42 | 2.38 | 2.34 | 2.31 | 2.23 | 2.16 | 2.11 | 2.07 | 2.03 | 1.98 | 1.93 |
| 20 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 | 2.35 | 2.31 | 2.28 | 2.20 | 2.12 | 2.08 | 2.04 | 1.99 | 1.95 | 1.90 |
| 21 | 4.32 | 3.47 | 3.07 | 2.84 | 2.68 | 2.57 | 2.49 | 2.42 | 2.37 | 2.32 | 2.28 | 2.25 | 2.18 | 2.10 | 2.05 | 2.01 | 1.96 | 1.92 | 1.87 |
| 22 | 4.30 | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.46 | 2.40 | 2.34 | 2.30 | 2.26 | 2.23 | 2.15 | 2.07 | 2.03 | 1.98 | 1.94 | 1.89 | 1.84 |
| 23 | 4.28 | 3.42 | 3.03 | 2.80 | 2.64 | 2.53 | 2.44 | 2.37 | 2.32 | 2.27 | 2.24 | 2.20 | 2.13 | 2.05 | 2.01 | 1.96 | 1.91 | 1.86 | 1.81 |
| 24 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 | 2.25 | 2.22 | 2.18 | 2.11 | 2.03 | 1.98 | 1.94 | 1.89 | 1.84 | 1.79 |
| 25 | 4.24 | 3.39 | 2.99 | 2.76 | 2.60 | 2.49 | 2.40 | 2.34 | 2.28 | 2.24 | 2.20 | 2.16 | 2.09 | 2.01 | 1.96 | 1.92 | 1.87 | 1.82 | 1.77 |
| 26 | 4.23 | 3.37 | 2.98 | 2.74 | 2.59 | 2.47 | 2.39 | 2.32 | 2.27 | 2.22 | 2.18 | 2.15 | 2.07 | 1.99 | 1.95 | 1.90 | 1.85 | 1.80 | 1.75 |
| 27 | 4.21 | 3.35 | 2.96 | 2.73 | 2.57 | 2.46 | 2.37 | 2.31 | 2.25 | 2.20 | 2.17 | 2.13 | 2.06 | 1.97 | 1.93 | 1.88 | 1.84 | 1.79 | 1.73 |
| 28 | 4.20 | 3.34 | 2.95 | 2.71 | 2.56 | 2.45 | 2.36 | 2.29 | 2.24 | 2.19 | 2.15 | 2.12 | 2.04 | 1.96 | 1.91 | 1.87 | 1.82 | 1.77 | 1.71 |
| 29 | 4.18 | 3.33 | 2.93 | 2.70 | 2.55 | 2.43 | 2.35 | 2.28 | 2.22 | 2.18 | 2.14 | 2.10 | 2.03 | 1.94 | 1.90 | 1.85 | 1.81 | 1.75 | 1.70 |
| 30 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 | 2.16 | 2.13 | 2.09 | 2.01 | 1.93 | 1.89 | 1.84 | 1.79 | 1.74 | 1.68 |
| 40 | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.25 | 2.18 | 2.12 | 2.08 | 2.04 | 2.00 | 1.92 | 1.84 | 1.79 | 1.74 | 1.69 | 1.64 | 1.58 |
| 60 | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.10 | 2.04 | 1.99 | 1.95 | 1.92 | 1.84 | 1.75 | 1.70 | 1.65 | 1.59 | 1.53 | 1.47 |
| 120 | 3.92 | 3.07 | 2.68 | 2.45 | 2.29 | 2.18 | 2.09 | 2.02 | 1.96 | 1.91 | 1.87 | 1.83 | 1.75 | 1.66 | 1.61 | 1.55 | 1.50 | 1.43 | 1.35 |

Propiedades de la distribución F.

i) Si $T \sim t_n$, entonces $T^2 \sim F_{1,n}$.

Demostración: En efecto, si $T \sim t_n$, entonces

$$T = \frac{N(0,1)}{\sqrt{\frac{\chi_n^2}{n}}}.$$

Por tanto,

$$T^2 = \frac{N(0,1)^2}{\frac{\chi_n^2}{n}} = \frac{\chi_1^2}{\frac{\chi_n^2}{n}} = \frac{\chi_1^2}{\frac{\chi_n^2}{n}} \sim F_{1,n}.$$

ii) Si $F \sim F_{m,n}$, entonces $\frac{1}{F} \sim F_{n,m}$.

Demostración: Ciertamente, pues si $F \sim F_{m,n} \implies F = \frac{\frac{\chi_m^2}{m}}{\frac{\chi_n^2}{n}}$, siendo numerador y denominador dos distribuciones ji-cuadradas independientes, entonces $\frac{1}{F} = \frac{\frac{\chi_n^2}{n}}{\frac{\chi_m^2}{m}}$. Esta propiedad es

muy útil para cálculos asociados a colas izquierdas de probabilidad en la distribución F .

Ejemplo 6.8: Calcular x tal que $P(F_{4,10} < x) = 0,05$.

$$\begin{aligned} 0,05 = P(F_{4,10} < x) &= P\left(\frac{1}{F_{4,10}} > \frac{1}{x}\right) = P(F_{10,4} > \frac{1}{x}) \\ \implies \frac{1}{x} &=_{\text{tabla}} 5,96 \implies x = \frac{1}{5,96}. \end{aligned}$$

El resultado teórico que hace tan relevantes las distribuciones que acabamos de introducir se conoce como Teorema de Fisher.

6.6 Teorema de Fisher.

Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de $N(\mu, \sigma)$. Entonces:

i)

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \implies \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1).$$

ii)

$$\frac{nS^2}{\sigma^2} = \frac{\sum (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2,$$

ó

$$\frac{(n-1)\hat{S}^2}{\sigma^2} = \frac{\sum (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2.$$

iii) \bar{X} y $\frac{nS^2}{\sigma^2}$ son dos variables aleatorias independientes.

6.6.1 Consecuencias del Teorema de Fisher.

i) Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de $N(\mu, \sigma)$. Entonces:

$$T = \frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{nS^2}{\frac{\sigma^2}{n-1}}}} \sim t_{n-1}.$$

Ya que es el cociente entre una normal tipificada y la raíz de una ji-cuadrado corregida por sus grados de libertad, ambas independientes. Simplificando,

$$T = \frac{\bar{X} - \mu}{S} \sqrt{n-1} \sim t_{n-1}.$$

Si se desea escribir el resultado en términos de la cuasidesviación típica, entonces:

$$T = \frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)\hat{S}^2}{\frac{\sigma^2}{n-1}}}} = \frac{\bar{X} - \mu}{\hat{S}} \sqrt{n} \sim t_{n-1}.$$

ii) Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de $N(\mu, \sigma)$ e Y_1, Y_2, \dots, Y_m otra muestra aleatoria simple también de $N(\mu', \sigma')$, ambas muestras independientes entre sí. Entonces:

$$F = \frac{\frac{nS_X^2}{\sigma^2(n-1)}}{\frac{mS_Y^2}{(\sigma')^2(m-1)}},$$

donde el numerador y denominador siguen unas distribuciones $\frac{\chi_{n-1}^2}{n-1}$ y $\frac{\chi_{m-1}^2}{m-1}$, respectivamente. Luego $F \rightsquigarrow F_{n-1, m-1}$. Simplificando,

$$F = \frac{n(m-1)S_X^2(\sigma')^2}{m(n-1)S_Y^2\sigma^2} \rightsquigarrow F_{n-1, m-1}.$$

En particular, si $\sigma = \sigma'$, entonces

$$F = \frac{n(m-1)S_X^2}{m(n-1)S_Y^2} \rightsquigarrow F_{n-1, m-1}.$$

Si escribimos el resultado en términos de la cuasivarianza se obtiene $\frac{\hat{S}_X^2}{\hat{S}_Y^2} \rightsquigarrow F_{n-1, m-1}$ dado que

$$\frac{nS_X^2}{n-1} = \hat{S}_X^2,$$

y

$$\frac{mS_Y^2}{m-1} = \hat{S}_Y^2.$$

6.7 El estadístico de orden:

Definición 6.7: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de una población X (variable aleatoria). Llamaremos estadísticos de orden o estadísticos ordenados a los estadísticos que ordenan la muestra de menor a mayor valor. Se notan $X_{(1)}, X_{(2)}, \dots, X_{(n)}$. Donde $X_{(1)} = \min(\{X_1, X_2, \dots, X_n\})$ y $X_{(n)} = \max(\{X_1, X_2, \dots, X_n\})$.

Ejemplo 6.9: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de una población X variable aleatoria. Si la realización muestral es; 3, 7, -2, 0, 3, 5.

Entonces,

$$\bar{X} = \frac{3+7-2+3+5}{3} = \frac{8}{3},$$

$$S^2 = 8,89,$$

$$\hat{S}^2 = 10,67,$$

y

$$X_{(1)}, X_{(2)}, \dots, X_{(n)} = (-2, 0, 3, 3, 5, 7).$$

Si $f(x)$ representa la densidad o función de probabilidad de X entonces

$$f(x_1, x_2, \dots, x_n),$$

será la densidad o función de probabilidad de la muestra aleatoria simple. Dada la independencia de la muestra se calcula como:

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i).$$

Y la del estadístico ordenado:

$$f(x_{(1)}, x_{(2)}, \dots, x_{(n)}) = n! f(x_1, x_2, \dots, x_n) = n! \prod_{i=1}^n f(x_i).$$

A la densidad o función de probabilidad de la muestra se la denomina la verosimilitud muestral, cuyo concepto se extenderá en el siguiente capítulo.

Ejemplo 6.10: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de una $U(0, 1)$. Entonces $X_{(1)} \rightsquigarrow B(1, n)$ y $X_{(n)} \rightsquigarrow B(n, 1)$. En efecto, como ni $X_{(1)}$ ni $X_{(n)}$ son transformaciones biyectivas de la muestra, no es utilizable el teorema jacobiano y debemos obtener la función de distribución de cada variable. Si $x \in (0, 1)$,

$$\begin{aligned} F_{X_{(1)}}(x) &= P(\min_{i=1, \dots, n} \{X_i\} \leq x) = 1 - P(\min_{i=1, \dots, n} \{X_i\} \geq x) \\ &= 1 - P(X_1 > x, \dots, X_n > x) = 1 - \prod_{i=1}^n P(X_i > x) \\ &= 1 - p(X_1 > x)^n = 1 - (1 - x)^n, \end{aligned}$$

luego

$$F_{X_{(1)}}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - (1 - x)^n & \text{si } x \in (0, 1) \\ 1 & \text{si } x \geq 1. \end{cases}$$

De donde

$$f_{X_{(1)}}(x) = \begin{cases} n(1 - x)^{n-1} & \text{si } x \in (0, 1) \\ 0 \text{ fuera.} & \end{cases} \rightsquigarrow B(1, n).$$

Análogamente si $x \in (0, 1)$

$$\begin{aligned} F_{X_{(n)}}(x) &= P(\max_{i=1, \dots, n} \{X_i\} \leq x) = P(X_1 \leq x, \dots, X_n \leq x) \\ &= \prod_{i=1}^n P(X_i \leq x) = P(X_1 \leq x)^n = x^n. \end{aligned}$$

Entonces,

$$F_{X_{(n)}}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x^n & \text{si } x \in (0, 1) \\ 1 & \text{si } x > 1, \end{cases}$$

y, por lo tanto

$$f_{X_{(n)}}(x) = \begin{cases} nx^{n-1} & \text{si } x \in (0, 1) \\ 0 \text{ fuera.} & . \end{cases} \rightsquigarrow B(n, 1).$$

Bibliografía

1. A.A. Borovkov (2013). *Probability Theory*. Springer.
2. A. Klenke (2014). *Probability Theory. A Comprehensive Course*. Springer.
3. M.H. DeGroot (1988). *Probability and Statistics*. Wilmington, Delaware, E.U.A. Addison-Wesley Iberoamericana, S.A.
4. V.K. Rohatgi and A.K. Md. Ehsanes Saleh (2001). *An Introduction to Probability and Statistics*. John Wiley and Sons.

Tema 7

Principios de reducción de datos. Introducción. El principio de suficiencia. El principio de verosimilitud: la función de verosimilitud y el principio formal de verosimilitud.

Supongamos, como hasta ahora, que nos encontramos en un modelo paramétrico, con una población X que sigue una función de probabilidad o de densidad $f(x, \theta)$, donde θ es un parámetro o vector de parámetros desconocidos y que $\theta \in \Theta$, el llamado espacio paramétrico. El objetivo es la estimación puntual de θ a partir de una muestra aleatoria simple X_1, X_2, \dots, X_n de una población X .

7.1 Reducción de datos. Introducción

Supóngase que un problema de estimación específico, dos científicos, C_1 y C_2 , deben estimar el valor del parámetro θ ; que el científico C_1 puede observar los valores de las observaciones $X_1 = x_1, \dots, X_n = x_n$ de una muestra aleatoria; y que el científico C_2 no puede observar los valores $X_1 = x_1, \dots, X_n = x_n$ pero puede saber el valor de cierto estadístico $T = T(X_1, \dots, X_n)$. En este caso, el científico C_1 puede elegir cualquier función de las observaciones $X_1 = x_1, \dots, X_n = x_n$ como un estimador de θ , mientras que el científico C_2 puede utilizar únicamente una función de T . Por tanto, resulta que C_1 generalmente podrá encontrar una mejor estimación que C_2 .

7.1.1 Principio de suficiencia.

En algunos problemas, sin embargo, C_2 podrá hacerlo tan bien como C_1 . En tal caso, la función $T = T(X_1, \dots, X_n)$ resumirá, en cierto sentido, toda la información contenida en la muestra aleatoria, y será irrelevante el conocimiento de los valores de X_1, \dots, X_n en la búsqueda de un buen estimador de θ . Un estadístico T que tiene esta propiedad se llama estadístico suficiente. Se presentará ahora la definición formal de estadístico suficiente.

Definición 7.1: Si T es un estadístico y t es un valor concreto de T , entonces la distribución conjunta condicional de X_1, \dots, X_n , dado que $T = t$, se puede calcular a partir de la distribución condicional. En general, esta distribución conjunta condicional dependerá del valor de θ . Por tanto, para cada valor de t , existirá una familia de distribuciones condicionales posibles que corresponden a los distintos valores posibles de $\theta \in \Theta$. Puede suceder, sin embargo, que para cada valor posible de t , tal distribución condicional sea la misma para todos los valores $\theta \in \Theta$ y, por tanto, realmente no depende del valor de θ . En este caso, se dice que T es un estadístico suficiente para el parámetro θ .

Antes de describir un método sencillo para encontrar estadísticos suficientes y antes de considerar ejemplos de estadísticos suficientes, se indicará por qué un estadístico suficiente T que satisface la definición que se acaba de presentar se considera como un resumen de toda la información relevante acerca de θ contenida en la muestra X_1, \dots, X_n . Considérese de nuevo el caso del científico

C_2 que sólo puede saber el valor del estadístico T y no puede observar los valores de X_1, \dots, X_n . Si T es un estadístico suficiente, entonces la distribución conjunta condicional de X_1, \dots, X_n dado que $T = t$, es completamente conocida para cualquier valor observado t y no depende del valor desconocido de θ . Por tanto, para cualquier valor t que se podría observar, el científico C_2 podría, en principio, generar n variables aleatorias X'_1, \dots, X'_n de acuerdo con esta distribución conjunta condicional. El proceso de generar variables aleatorias X'_1, \dots, X'_n que tienen una distribución de probabilidad conjunta específica se llama una aleatorización auxiliar.

Cuando utilizamos este proceso de observar primero T y luego generar X'_1, \dots, X'_n de acuerdo con la distribución conjunta condicional específica, resulta que para cualquier valor concreto de $\theta \in \Theta$, la distribución conjunta marginal de X'_1, \dots, X'_n será la misma que la distribución conjunta de X_1, \dots, X_n . Por tanto, si el científico C_2 puede observar el valor de un estadístico suficiente T , entonces puede generar n variables aleatorias X'_1, \dots, X'_n que tenga la misma distribución conjunta que la muestra aleatoria original X_1, \dots, X_n . La propiedad que distingue un estadístico suficiente T de un estadístico que no es suficiente se puede describir como sigue: la aleatorización auxiliar utilizada para generar las variables aleatorias X'_1, \dots, X'_n después de que ha sido observado el estadístico no requiere ningún conocimiento acerca del valor de θ , puesto que la distribución conjunta condicional de X_1, \dots, X_n dado el valor de T no depende del valor de θ . Si el estadístico T no fuera suficiente, esta aleatorización auxiliar no podría llevarse a cabo, porque la distribución conjunta condicional de X_1, \dots, X_n para un valor dado de T involucraría el valor de θ y este valor es desconocido.

Se puede demostrar ahora por qué el científico C_2 , que puede observar únicamente el valor del estadístico suficiente T , no obstante, puede estimar θ tan bien como puede hacerlo el científico C_1 , que observa los valores de X_1, \dots, X_n . Supóngase que C_1 decide utilizar un estimador concreto $U(X_1, \dots, X_n)$ para estimar θ y que C_2 observa el valor de T y genera las variables X'_1, \dots, X'_n que tienen la misma distribución conjunta de X_1, \dots, X_n . Si C_2 utiliza el estimador $U(X'_1, \dots, X'_n)$, entonces resulta que la distribución de probabilidad del estimador C_2 será la misma que la distribución de probabilidad del estimador de C_1 . Esta exposición ilustra por qué, cuando se busca un buen estimador, un científico puede restringir la búsqueda a estimadores que son funciones de un estadístico suficiente T .

7.1.2 Criterio de factorización

Se presenta ahora un método sencillo para obtener un estadístico suficiente que se puede aplicar a muchos problemas. Este método está basado en el siguiente resultado, que fue desarrollado cada vez con más generalidad por R.A. Fisher en 1922, J. Neyman en 1935 y P.R. Halmos y L. J. Savage en 1949.

Criterio de factorización. Sea X_1, \dots, X_n una muestra aleatoria de una distribución continua o una distribución discreta cuya función de probabilidad o función de densidad es $f(x, \theta)$, donde el valor θ es desconocido y pertenece a un espacio paramétrico Θ concreto. Un estadístico $T = T(X_1, \dots, X_n)$ es un estadístico suficiente para θ si, y sólo si, la función de probabilidad o la función de densidad conjunta $f(x_1, \dots, x_n, \theta)$ se puede factorizar como sigue para todos los

valores de $x_1, \dots, x_n \in \mathbb{R}^n$ y todos los valores de $\theta \in \Theta$:

$$f(x_1, \dots, x_n, \theta) = h(x_1, \dots, x_n)g(x_1, \dots, x_n, \theta). \quad (7.1)$$

Aquí, las funciones h y g son no negativas; la función h puede depender de X_1, \dots, X_n pero no de θ y la función g dependerá de θ pero depende de los valores observados X_1, \dots, X_n únicamente a través del valor del estadístico $T(X_1, \dots, X_n)$.

Demostración: se dará la demostración únicamente para el caso en que las variables aleatorias X_1, \dots, X_n tienen una distribución discreta, en cuyo caso

$$f(x_1, \dots, x_n, \theta) = P(X_1 = x_1, \dots, X_n = x_n, \theta).$$

Supóngase en primer lugar que $f(x_1, \dots, x_n, \theta)$ se puede factorizar como en 7.1 para todos los valores de $x_1, \dots, x_n \in \mathbb{R}^n$ y $\theta \in \Theta$. Para cada valor posible t de T , sea $A(t)$ el conjunto de todos los puntos $x_1, \dots, x_n \in \mathbb{R}^n$ tales que $T(x_1, \dots, x_n) = t$. Para cualquier valor concreto de $\theta \in \Theta$, se determinará la distribución condicional de X_1, \dots, X_n dado que $T = t$. Para cualquier punto $x_1, \dots, x_n \in A(t)$,

$$P(X_1 = x_1, \dots, X_n = x_n | T = t, \theta) = \frac{P(X_1 = x_1, \dots, X_n = x_n, \theta)}{P(T = t, \theta)} = \frac{f(x_1, \dots, x_n, \theta)}{\sum_{y_1, \dots, y_n \in A(t)} f(y_1, \dots, y_n, \theta)}.$$

Puesto que $T(y_1, \dots, y_n) = t$ para todo $y_1, \dots, y_n \in A(t)$, y puesto que $x_1, \dots, x_n \in A(t)$ de la ecuación 7.1 resulta que

$$P(X_1 = x_1, \dots, X_n = x_n | T = t, \theta) = \frac{h(x_1, \dots, x_n)}{\sum_{y_1, \dots, y_n \in A(t)} g(y_1, \dots, y_n, \theta)} \quad (7.2)$$

Por último, para cualquier punto x_1, \dots, x_n que no pertenece a $A(t)$,

$$P(X_1 = x_1, \dots, X_n = x_n | T = t, \theta) = 0. \quad (7.3)$$

De las ecuaciones 7.2 y 7.3 se puede ver que la distribución condicional de X_1, \dots, X_n no depende de θ . Por tanto, T es un estadístico suficiente.

Recíprocamente, supongamos que T es un estadístico suficiente. Entonces, para cualquier valor concreto t de T , cualquier punto $x_1, \dots, x_n \in A(t)$ y cualquier valor de $\theta \in \Theta$, la probabilidad condicional $P(X_1 = x_1, \dots, X_n = x_n | T = t, \theta)$ no dependerá de θ y por tanto tendrá la forma

$$P(X_1 = x_1, \dots, X_n = x_n | T = t, \theta) = h(x_1, \dots, x_n).$$

Si se define $g(t, \theta) = P(T = t, \theta)$, resulta que

$$f(x_1, \dots, x_n, \theta) = P(X_1 = x_1, \dots, X_n = x_n, \theta) = P(X_1 = x_1, \dots, X_n = x_n | T = t, \theta)P(T = t, \theta)$$

$$= h(x_1, \dots, x_n)g(t, \theta).$$

Por tanto, $f(x_1, \dots, x_n, \theta)$ se ha factorizado en la forma dada por la ecuación 7.1.

La demostración para una muestra aleatoria X_1, \dots, X_n de una distribución continua requiere métodos distintos y no se proporcionará aquí.

Para cualquier valor de x_1, \dots, x_n cuya $f(x_1, \dots, x_n, \theta) = 0$ para todos los valores de $\theta \in \Theta$, el valor de la función $h(x_1, \dots, x_n)$ de la ecuación 7.1 se puede elegir como 0. Por tanto, cuando se aplica el criterio de factorización, es suficiente comprobar que una factorización de la forma dada por la ecuación 7.1 se satisface para todo valor de x_1, \dots, x_n tal que $P(x_1, \dots, x_n, \theta) > 0$ para al menos un valor de $\theta \in \Theta$.

Se ilustrará ahora el uso del criterio de factorización en el siguiente ejemplo.

Ejemplo 7.1: Sea X_1, \dots, X_n una muestra aleatoria simple de una población X . Determinar un estadístico suficiente para estimar el parámetro desconocido en cada uno de los siguientes casos:

- a) p si $X \sim \text{Bin}(m, p)$, p desconocido.
- b) θ si $X \sim U(0, \theta)$, θ desconocido.
- c) λ si $X \sim \exp(\lambda)$, λ desconocido.
- d) a si $X \sim \Gamma(a, p_0)$, a desconocido.
- e) p si $X \sim \Gamma(a_0, p)$, p desconocido.
- f) μ si $X \sim N(\mu, \sigma_0)$, μ desconocido.
- g) σ si $X \sim N(\mu_0, \sigma)$, σ desconocido.
- h) p si $X \sim P(\lambda)$, λ desconocido.

- a) p si $X \sim \text{Bin}(m, p)$.

$$\begin{aligned} f(x_1, x_2, \dots, x_n, p) &= \prod_{i=1}^n f(x_i, p) = \prod_{i=1}^n \left[\binom{m}{x_i} p^{x_i} (1-p)^{m-x_i} \right] \\ &= \begin{cases} \left[\prod_{i=1}^n \binom{m}{x_i} \right] p^{\sum_{i=1}^n x_i} (1-p)^{nm - \sum_{i=1}^n x_i} & \text{si } x_1, x_2, \dots, x_n \in \{0, 1, 2, \dots, m\} \\ 0 & \text{fuera.} \end{cases} \\ &= \underbrace{\left[\prod_{i=1}^n \binom{m}{x_i} \right] \cdot h_1(x_1, x_2, \dots, x_n)}_{h(x_1, x_2, \dots, x_n)} \cdot \underbrace{p^{\sum_{i=1}^n x_i} (1-p)^{nm - \sum_{i=1}^n x_i}}_{g(t, p)}, \end{aligned}$$

siendo

$$h_1(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, x_2, \dots, x_n \in \{0, 1, 2, \dots, m\} \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \sum_{i=1}^n X_i$ es un estadístico suficiente para estimar p .

b) θ si $X \sim U(0, \theta)$.

$$\begin{aligned} f(x_1, x_2, \dots, x_n, \theta) &= \prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n \frac{1}{\theta} = \begin{cases} \frac{1}{\theta^n} & \text{si } x_1, x_2, \dots, x_n \in (0, \theta) \\ 0 & \text{fuera.} \end{cases} \\ &= \begin{cases} \frac{1}{\theta^n} & \text{si } \min\{x_1, x_2, \dots, x_n\} > 0 \text{ y } \max\{x_1, x_2, \dots, x_n\} < \theta \\ 0 & \text{fuera.} \end{cases} \\ &= \frac{1}{\theta^n} \cdot h_1(x_1, x_2, \dots, x_n) g_1(x_1, x_2, \dots, x_n, \theta), \end{aligned}$$

siendo

$$h_1(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } \min\{x_1, x_2, \dots, x_n\} > 0 \\ 0 & \text{fuera.} \end{cases}$$

y

$$g_1(x_1, x_2, \dots, x_n, \theta) = \begin{cases} 1 & \text{si } \max\{x_1, x_2, \dots, x_n\} < \theta \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \max\{X_1, X_2, \dots, X_n\}$ es un estadístico suficiente para estimar θ .

c) λ si $X \sim \exp(\lambda)$.

$$\begin{aligned} f(x_1, x_2, \dots, x_n, \lambda) &= \prod_{i=1}^n f(x_i, \lambda) = \prod_{i=1}^n (\lambda e^{-\lambda x_i}) \\ &= \begin{cases} \lambda^n e^{-\lambda \sum_{i=1}^n x_i} & \text{si } x_1, x_2, \dots, x_n \in \mathbb{R}^+ \\ 0 & \text{fuera.} \end{cases} \\ &= \underbrace{\lambda^n e^{-\lambda \sum_{i=1}^n x_i}}_{g(t=\sum_{i=1}^n x_i, \lambda)} \cdot h_1(x_1, x_2, \dots, x_n), \end{aligned}$$

siendo

$$h_1(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, x_2, \dots, x_n \in \mathbb{R}^+ \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \sum_{i=1}^n X_i$ es un estadístico suficiente para estimar λ .

d) a si $X \sim \Gamma(a, p_0)$, p_0 conocido.

$$f(x_1, x_2, \dots, x_n, a) = \prod_{i=1}^n f(x_i, a) = \prod_{i=1}^n \left[\frac{p_0^a}{\Gamma(a)} e^{-p_0 x_i} x_i^{a-1} \right]$$

$$\begin{aligned}
&= \begin{cases} \frac{p_0^{na}}{[\Gamma(a)]^n} e^{-p_0 \sum_{i=1}^n x_i} (\prod_{i=1}^n x_i)^{a-1} & \text{si } x_1, x_2, \dots, x_n \in \mathbb{R}^+ \\ 0 & \text{fuera.} \end{cases} \\
&= \underbrace{\frac{p_0^{na}}{[\Gamma(a)]^n} (\prod_{i=1}^n x_i)^{a-1}}_{g(t=\prod_{i=1}^n x_i, a)} \underbrace{e^{-p_0 \sum_{i=1}^n x_i} \cdot h_1(x_1, x_2, \dots, x_n)}_{h(x_1, x_2, \dots, x_n)},
\end{aligned}$$

siendo

$$h_1(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, x_2, \dots, x_n \in \mathbb{R}^+ \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \prod_{i=1}^n X_i$ es un estadístico suficiente para estimar a .

e) p si $X \sim \Gamma(a_0, p)$, a_0 conocido.

$$\begin{aligned}
f(x_1, x_2, \dots, x_n, p) &= \prod_{i=1}^n f(x_i, p) = \prod_{i=1}^n \left[\frac{p^{a_0}}{\Gamma(a_0)} e^{-px_i} x_i^{a_0-1} \right] \\
&= \begin{cases} \frac{p^{na_0}}{[\Gamma(a_0)]^n} e^{-p \sum_{i=1}^n x_i} (\prod_{i=1}^n x_i)^{a_0-1} & \text{si } x_1, x_2, \dots, x_n \in \mathbb{R}^+ \\ 0 & \text{fuera.} \end{cases} \\
&= \underbrace{p^{na_0} e^{-p \sum_{i=1}^n x_i}}_{g(t=\sum_{i=1}^n x_i, p)} \underbrace{\frac{(\prod_{i=1}^n x_i)^{a_0-1}}{[\Gamma(a_0)]^n} \cdot h_1(x_1, x_2, \dots, x_n)}_{h(x_1, x_2, \dots, x_n)},
\end{aligned}$$

siendo

$$h_1(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, x_2, \dots, x_n \in \mathbb{R}^+ \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \sum_{i=1}^n X_i$ es un estadístico suficiente para estimar p .

f) μ si $X \sim N(\mu, \sigma_0)$, σ_0 conocido.

$$\begin{aligned}
f(x_1, x_2, \dots, x_n, \mu) &= \prod_{i=1}^n f(x_i, \mu) = \prod_{i=1}^n \left[\frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma_0^2}} \right] \\
&= \frac{1}{(\sigma_0 \sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma_0^2}} \\
&= \underbrace{\frac{1}{(\sigma_0 \sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i)^2}{2\sigma_0^2}}}_{h(x_1, x_2, \dots, x_n)} \cdot \underbrace{e^{-\frac{\mu^2 n}{2\sigma_0^2}} e^{\frac{\sum_{i=1}^n x_i \mu}{\sigma_0^2}}}_{g(t=\sum_{i=1}^n x_i, \mu)}
\end{aligned}$$

Por el criterio de factorización $T = \sum_{i=1}^n X_i$ es el estadístico suficiente para estimar μ .

g) σ si $X \rightsquigarrow N(\mu_0, \sigma)$, μ_0 conocido.

$$\begin{aligned} f(x_1, x_2, \dots, x_n, \sigma) &= \prod_{i=1}^n f(x_i, \sigma) = \prod_{i=1}^n \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - \mu_0)^2}{2\sigma^2}} \right] \\ &= \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i - \mu_0)^2}{2\sigma^2}} \\ &= \underbrace{\frac{1}{(\sqrt{2\pi})^n}}_{h(x_1, x_2, \dots, x_n)} \cdot \underbrace{\frac{1}{\sigma^n} e^{-\frac{\sum_{i=1}^n (x_i - \mu_0)^2}{2\sigma^2}}}_{g(t = \sum_{i=1}^n (x_i - \mu_0)^2)} \end{aligned}$$

Por el criterio de factorización $T = \sum_{i=1}^n (X_i - \mu_0)^2$ es el estadístico suficiente para estimar σ .

h) λ si $X \rightsquigarrow P(\lambda)$.

$$\begin{aligned} f(x_1, x_2, \dots, x_n, \lambda) &= \prod_{i=1}^n f(x_i, \lambda) = \prod_{i=1}^n \left[\frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right] \\ &= \begin{cases} e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} & \text{si } x_1, x_2, \dots, x_n \in \mathbb{N} \cup \{0\} \\ 0 & \text{fuera.} \end{cases} \\ &= \underbrace{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}_{g(t = \sum_{i=1}^n x_i, \lambda)} \cdot \underbrace{\frac{1}{\prod_{i=1}^n x_i!} \cdot h_1(x_1, x_2, \dots, x_n)}_{h(x_1, x_2, \dots, x_n)} \end{aligned}$$

siendo

$$h_1(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, x_2, \dots, x_n \in \mathbb{N} \cup \{0\} \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \sum_{i=1}^n X_i$ es el estadístico suficiente para estimar λ .

7.2 Estadísticos conjuntamente suficientes

Se continuará suponiendo que las variables X_1, \dots, X_n constituyen una muestra aleatoria de una distribución cuya función de probabilidad o función de densidad es $f(x_1, \dots, x_n, \theta)$, donde el parámetro θ debe pertenecer a un espacio paramétrico Θ . Sin embargo, se considera explícitamente la posibilidad de que θ pueda ser un vector de parámetros reales. Por ejemplo, si la muestra proviene de una distribución normal donde la media μ y la varianza σ^2 son desconocidas, entonces θ sería un vector bidimensional cuyas componentes son μ y σ^2 . Análogamente, si la muestra

proviene de una distribución uniforme sobre el intervalo (a, b) donde ambos puntos extremos, a y b , son desconocidos, entonces θ sería un vector bidimensional cuyas componentes son a y b . Se continuará, por supuesto, incluyendo la posibilidad de que θ sea un parámetro unidimensional.

En casi todo problema en el que θ es un vector, al igual que en muchos problemas en los que θ es unidimensional, no existe un único estadístico T que sea suficiente. En tal caso es necesario encontrar dos o más estadísticos T_1, \dots, T_k que son estadísticos conjuntamente suficientes en un sentido que se describirá ahora.

Supóngase que en un problema concreto los estadísticos T_1, \dots, T_k se definen por k funciones distintas de las observaciones X_1, \dots, X_n . Específicamente, sea $T_i = T_i(X_1, \dots, X_n)$ para $i = 1, \dots, k$. En términos generales, los estadísticos T_1, \dots, T_k son estadísticos conjuntamente suficientes para θ si un investigador que asume únicamente los valores de las k funciones $T_1(X_1, \dots, X_n), \dots, T_k(X_1, \dots, X_n)$ puede estimar cualquier componente de θ , o cualquier función de los componentes de θ , tan bien como puede hacerlo un investigador que observa los n valores de X_1, \dots, X_n . Desde el punto de vista del criterio de factorización, se puede formular la siguiente versión.

Los estadísticos T_1, \dots, T_k son estadísticos conjuntamente suficientes para θ si, y sólo si, la función de probabilidad conjunta o la función de densidad conjunta $f(x_1, \dots, x_n, \theta)$ se puede factorizar como sigue para todos los valores de $x_1, \dots, x_n \in \mathbb{R}^n$ y todos los valores de $\theta \in \Theta$:

$$f(x_1, \dots, x_n, \theta) = h(x_1, \dots, x_n)g[T_1(x_1, \dots, x_n), \dots, T_k(x_1, \dots, x_n), \theta]. \quad (7.4)$$

Aquí las funciones h y g son no negativas, la función h puede depender de x_1, \dots, x_n , pero no depende de θ , y la función g dependerá de θ , pero dependerá de x_1, \dots, x_n únicamente a través de las k funciones $T_1(x_1, \dots, x_n), \dots, T_k(x_1, \dots, x_n)$.

Ejemplo 7.2:

Estadísticos conjuntamente suficientes para los parámetros de una distribución normal. Supóngase que X_1, \dots, X_n constituyen una muestra aleatoria simple de una distribución con media μ y varianza σ^2 desconocidas. La función de densidad conjunta de X_1, \dots, X_n esta dada por la ecuación

$$\begin{aligned} f(x_1, \dots, x_n, \mu) &= \prod_{i=1}^n \left[\frac{1}{(2\pi)^{\frac{n}{2}} \sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \right] \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp\left[-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n x_i^2 - \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i + \frac{n\mu^2}{2\sigma^2} \right] \right]. \end{aligned}$$

Se puede observar que la función de densidad conjunta depende de x_1, \dots, x_n a través de los valores de $\sum_{i=1}^n x_i$ y $\sum_{i=1}^n x_i^2$. Por tanto, por el criterio de factorización, los estadísticos $T_1 = \sum_{i=1}^n X_i$ y

$T_2 = \sum_{i=1}^n X_i^2$ son estadísticos conjuntamente suficientes para μ y σ .

Supóngase ahora que en un problema concreto los estadísticos T_1, \dots, T_k son estadísticos conjuntamente suficientes para un vector de parámetros θ . Si otros k estadísticos T'_1, \dots, T'_k se obtienen a partir de T_1, \dots, T_k mediante una transformación biunívoca, entonces se puede demostrar que T'_1, \dots, T'_k también serán estadísticos conjuntamente suficientes para θ .

Ejemplo 7.3:

Otro par de estadísticos conjuntamente suficientes para los parámetros de una distribución normal. Supóngase, de nuevo, que X_1, \dots, X_n constituyen una muestra aleatoria simple de una distribución normal con media μ y varianza σ^2 desconocidas y sean T'_1 y T'_2 la media muestral y la varianza muestral, respectivamente. Entonces,

$$T'_1 = \bar{X}_n,$$

y

$$T'_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Se demostrará ahora que T'_1 y T'_2 son estadísticos suficientes para μ y σ^2 .

Sean T_1 y T_2 los estadísticos conjuntamente suficientes para μ y σ^2 obtenidos en el ejemplo anterior. Entonces

$$T'_1 = \frac{1}{n} T_1,$$

y

$$T'_2 = \frac{1}{n} T_2 - \frac{1}{n^2} T_1^2.$$

Además, equivalentemente,

$$T_1 = nT'_1,$$

y

$$T_2 = n(T'_2 + T_1'^2).$$

Por tanto, los estadísticos T'_1 y T'_2 se obtienen a partir de los estadísticos conjuntamente suficientes T_1 y T_2 por medio de una transformación biunívoca. Resulta, por tanto, que T'_1 y T'_2 son estadísticos conjuntamente suficientes para μ y σ^2 .

Se ha demostrado ahora que los estadísticos conjuntamente suficientes para la media y varianza desconocidas de una distribución normal se pueden elegir como T_1 y T_2 , o T'_1 y T'_2 como se indica en los ejemplos anteriores.

Ejemplo 7.4: Estadísticos conjuntamente suficientes para los parámetros de una distribución uniforme. Supóngase que X_1, \dots, X_n constituyen una muestra aleatoria simple de una distribución uniforme sobre el intervalo $[a, b]$, donde los valores de ambos puntos extremos, a y b , son desconocidos, tales que $a < b$. Además, para todo x_1, \dots, x_n la función de densidad conjunta $f(x_1, \dots, x_n, a, b)$ de X_1, \dots, X_n será 0, a menos que todos los valores observados x_1, \dots, x_n

se encuentren entre a y b ; esto es, $f(x_1, \dots, x_n, a, b) = 0$, a menos que $\min\{x_1, \dots, x_n\} \geq a$ y $\max\{x_1, \dots, x_n\} \leq b$. Además, para todo x_1, \dots, x_n tal que $\min\{x_1, \dots, x_n\} \geq a$ y $\max\{x_1, \dots, x_n\} \leq b$ resulta que

$$f(x_1, \dots, x_n, a, b) = \frac{1}{(b-a)^n}.$$

Entonces,

$$f(x_1, \dots, x_n, a, b) = \frac{1}{(b-a)^n} h_1(x_1, \dots, x_n, a) h_2(x_1, \dots, x_n, b),$$

donde

$$h_1(x_1, \dots, x_n, a) = \begin{cases} 1 & \text{si } \min\{x_1, \dots, x_n\} \geq a \\ 0 & \text{en otro caso,} \end{cases}$$

y

$$h_2(x_1, \dots, x_n, b) = \begin{cases} 1 & \text{si } \max\{x_1, \dots, x_n\} \leq b \\ 0 & \text{en otro caso.} \end{cases}$$

Puesto que esta expresión depende de x_1, \dots, x_n únicamente a través de los valores de las funciones $\min\{x_1, \dots, x_n\}$ y $\max\{x_1, \dots, x_n\}$, resulta que los estadísticos $T_1 = \min\{X_1, \dots, X_n\}$ y $T_2 = \max\{X_1, \dots, X_n\}$ son estadísticos conjuntamente suficientes para a y b .

Ejemplo 7.5: Sea X_1, \dots, X_n muestra aleatoria simple de $X \sim \Gamma(a, p)$. Entonces

$$\begin{aligned} f(x_1, \dots, x_n, a, p) &= \prod_{i=1}^n f(x_i, a, p) = \begin{cases} \prod_{i=1}^n \left(\frac{p^a}{\Gamma(a)} e^{-px_i} x_i^{a-1} \right) & \text{si } x_1, \dots, x_n \geq 0 \\ 0 & \text{fuera} \end{cases} \\ &= \frac{p^{na}}{[\Gamma(a)]^n} e^{-p \sum_{i=1}^n x_i} \left(\prod_{i=1}^n x_i \right)^{a-1} h(x_1, \dots, x_n), \end{aligned}$$

con $h(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, \dots, x_n \geq 0 \\ 0 & \text{fuera.} \end{cases}$ Entonces $T_1 = \sum_{i=1}^n X_i$ y $T_2 = \prod_{i=1}^n X_i$ son estadísticos conjuntamente suficientes para estimar a y p .

Ejemplo 7.6: Sea X_1, \dots, X_n muestra aleatoria simple de $X \sim B(a, b)$. Entonces

$$\begin{aligned} f(x_1, \dots, x_n, a, b) &= \prod_{i=1}^n f(x_i, a, b) = \begin{cases} \prod_{i=1}^n \left[\frac{1}{\beta(a, b)} x_i^{a-1} (1-x_i)^{b-1} \right] & \text{si } x_1, \dots, x_n \in (0, 1) \\ 0 & \text{fuera} \end{cases} \\ &= \left(\frac{1}{\beta(a, b)} \right)^n \left(\prod_{i=1}^n x_i \right)^{a-1} \left[\prod_{i=1}^n (1-x_i) \right]^{b-1} h(x_1, \dots, x_n), \end{aligned}$$

con $h(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, \dots, x_n \in (0, 1) \\ 0 & \text{fuera.} \end{cases}$ Entonces $T_1 = \prod_{i=1}^n X_i$ y $T_2 = \prod_{i=1}^n (1-X_i)$ son estadísticos conjuntamente suficientes para estimar a y p .

7.2.1 Estadísticos suficientes minimales

En un problema concreto se trata de obtener un estadístico suficiente, o un conjunto de estadísticos conjuntamente suficientes para θ , porque los valores de tales estadísticos resumen toda la información relevante acerca de θ contenida en la muestra aleatoria. Cuando se conoce un conjunto de estadísticos conjuntamente suficientes, se simplifica la búsqueda de un buen estimador de θ porque sólo es necesario considerar funciones de estos estadísticos como posibles estimadores. Por tanto, en un problema concreto es conveniente encontrar no sólo un conjunto de estadísticos conjuntamente suficientes, sino el conjunto de estadísticos conjuntamente suficientes más sencillo. Por ejemplo, es correcto pero completamente inútil decir que en todo problema las n observaciones X_1, \dots, X_n son estadísticos conjuntamente suficientes.

Se describirá ahora otro conjunto de estadísticos conjuntamente suficientes que existe en todo problema y que es un poco más útil. Supóngase que X_1, \dots, X_n constituyen una muestra aleatoria simple de una distribución. Sea Y_1 el valor más pequeño de la muestra aleatoria, sea Y_2 el siguiente valor más pequeño, sea Y_3 el tercero más pequeño y así sucesivamente. De esta forma, Y_n , es el valor más grande de la muestra e Y_{n-1} es el valor más grande de los restantes. Las variables aleatorias Y_1, \dots, Y_n , como sabemos, se llaman estadísticos de orden de la muestra y se notan con frecuencia $(X_{(1)}, \dots, X_{(n)})$.

Ahora se definen $y_1 \leq y_2 \leq \dots \leq y_n$ como los valores de los estadísticos de orden de una muestra concreta. Si se proporcionan los valores y_1, \dots, y_n , entonces se sabe que esos n valores fueron obtenidos de la muestra. Sin embargo, no se sabe cuál de las observaciones X_1, \dots, X_n proporcionó realmente el valor de y_1 , cuál proporcionó realmente el valor de y_2 y así sucesivamente. Todo lo que se sabe es que el menor de los valores de X_1, \dots, X_n fue y_1 , y que el más pequeño de los restantes fue y_2 y así sucesivamente.

Si las variables X_1, \dots, X_n constituyen una muestra aleatoria de una distribución cuya función de probabilidad o función de densidad es $f(x_1, \dots, x_n, \theta)$, entonces la función de densidad conjunta o función de probabilidad conjunta de X_1, \dots, X_n tiene la siguiente forma:

$$f(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) \quad (7.5)$$

dada la independencia de las variables de la muestra. Puesto que el orden de los factores en el producto de la parte derecha de la ecuación 7.5 es irrelevante, la ecuación 7.5 podría escribirse también de la forma

$$f(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(y_i, \theta).$$

Por tanto, $f(x_1, \dots, x_n, \theta)$ depende de x_1, \dots, x_n únicamente a través de los valores de y_1, \dots, y_n . Resulta, por tanto, que los estadísticos de orden Y_1, \dots, Y_n siempre son estadísticos conjuntamente suficientes para θ . En otras palabras, es suficiente conocer el conjunto de n números que se obtiene de la muestra y no es necesario saber cuál de estos números en particular fue, por ejemplo, el valor de X_3 .

En cada uno de los ejemplos vistos en este tema, se consideró que existía un estadístico suficiente o dos estadísticos que eran conjuntamente suficientes. Para algunas distribuciones, sin embargo, los estadísticos de orden Y_1, \dots, Y_n constituyen el conjunto más sencillo que existe de estadísticos conjuntamente suficientes y no es posible una mayor reducción para estadísticos suficientes.

Ejemplo 7.7: Estadísticos suficientes para el parámetro de una distribución de Cauchy. Supóngase que X_1, \dots, X_n constituyen una muestra aleatoria simple de una distribución de Cauchy centrada en un punto desconocido θ , tal que $(-\infty < \theta < \infty)$. La función de densidad $f(x_1, \dots, x_n, \theta)$ está dada por

$$f(x, \theta) = \begin{cases} \frac{1}{\pi[1+(x_i-\theta)^2]} & \text{para } -\infty < x < \infty \\ 0 & \text{fuera,} \end{cases}$$

y la función de densidad conjunta $f(x_1, \dots, x_n, \theta)$ de X_1, \dots, X_n está dada por

$$f(x_1, \dots, x_n, \theta) = \begin{cases} \frac{1}{\pi^n \prod_{i=1}^n [1+(x_i-\theta)^2]} & \text{para } -\infty < x_i < \infty \text{ para todo } i = 1, \dots, n \\ 0 & \text{fuera.} \end{cases}$$

Se puede demostrar que los únicos estadísticos suficientes que existen en este problema son los estadísticos de orden Y_1, \dots, Y_n u otro conjunto de n estadísticos T_1, \dots, T_n que se pueden obtener a partir de los estadísticos de orden por medio de una transformación biunívoca. Los detalles del argumento no se darán aquí.

Estas consideraciones conducen a los conceptos de estadísticos minimales suficientes y a los conjuntos minimales de estadísticos conjuntamente suficientes. Hablando en términos generales, un estadístico T es un estadístico minimal suficiente si no se le puede reducir más sin violar la propiedad de suficiencia. Alternativamente, un estadístico suficiente T es minimal suficiente si toda función de T , que es en sí mismo un estadístico suficiente, es una función biunívoca de T . Formalmente, se utilizará la siguiente definición, que es equivalente a las definiciones informales que se acaban de dar.

Definición 7.2: Un estadístico T es un estadístico minimal suficiente si T es un estadístico suficiente y es una función de cualquier otro estadístico suficiente.

En cualquier problema donde existan estadísticos conjuntamente suficientes, los estadísticos minimales conjuntamente suficientes se definen de forma análoga.

7.3 Función de verosimilitud

La función de verosimilitud de una muestra aleatoria simple X_1, \dots, X_n de una población $X \sim f(x, \theta)$ se obtiene, como se ha dicho, dada la independencia de las variables de la muestra, como el producto de las funciones de densidad o probabilidad marginales, es decir,

$$f(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta).$$

Supóngase que dos conjuntos distintos de valores observados x_1, \dots, x_n e y_1, \dots, y_n que se podrían obtener del mismo experimento o de dos experimentos distintos tienen la propiedad de que determinan la misma función de verosimilitud para un cierto parámetro θ o determinan funciones de verosimilitud que son proporcionales entre sí. Entonces, x_1, \dots, x_n e y_1, \dots, y_n proporcionan la misma información acerca del valor desconocido de θ y un científico obtendrá la misma estimación de θ a partir de x_1, \dots, x_n o de y_1, \dots, y_n . A esta afirmación la conocemos como el principio de verosimilitud.

Por ejemplo, supóngase que un científico debe estimar la proporción desconocida θ de artículos defectuosos de un gran lote manufacturado. Supóngase, además, que se informa al científico que diez artículos fueron seleccionados al azar del lote y que resultaron exactamente dos defectuosos y ocho no defectuosos. Supóngase, sin embargo, que el científico no sabe cuál de los dos siguientes experimentos se ha llevado a cabo: (1) Una muestra fija de diez artículos ha sido seleccionada del lote y se encontró que dos de los artículos fueron defectuosos. (2) Se han seleccionado artículos al azar secuencialmente del lote hasta haber obtenido dos artículos defectuosos y se encontró que hubo que seleccionar un total de diez artículos.

Para cada uno de estos dos experimentos posibles, los valores observados determinan una función de verosimilitud que es proporcional a $\theta^2(1 - \theta)^8$ para $0 \leq \theta \leq 1$. Por tanto, si el científico utiliza un método de estimación que es compatible con el principio de verosimilitud, no necesita saber cuál de los dos experimentos posibles fue realizado realmente. Su estimación de θ sería igual en cualquier caso.

A continuación se calcula la función de verosimilitud para algunas de las distribuciones teóricas mas relevantes

Ejemplo 7.8:: Obtener la función de verosimilitud muestral si X_1, \dots, X_n es m.a.s. de X tal que:

- a) $X \rightsquigarrow B(1, p)$.
- b) $X \rightsquigarrow \text{Bin}(m, p)$.
- c) $X \rightsquigarrow P(\lambda)$.
- d) $X \rightsquigarrow U(a, b)$.
- e) $X \rightsquigarrow \exp(\lambda)$.
- f) $X \rightsquigarrow \Gamma(a, p)$.
- g) $X \rightsquigarrow N(\mu, \sigma)$.
- g) $X \rightsquigarrow B(a, b)$.

a) Si $X \rightsquigarrow B(1, p)$, entonces

$$f(x_1, \dots, x_n, p) = \prod_{i=1}^n f(x_i, p) = \prod_{i=1}^n [p^{x_i} (1-p)^{1-x_i}]$$

$$= \begin{cases} p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} & \text{si } x_i \in \{0, 1\}, \forall i = 1, 2, \dots, n \\ 0 & \text{fuera} \end{cases}$$

$$= p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} h(x_1, \dots, x_n),$$

siendo $h(x_1, \dots, x_n)$ la función indicador dada por $h(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, \dots, x_n \in \{0, 1\} \\ 0 & \text{fuera} \end{cases}$

b) Si $X \rightsquigarrow \text{Bin}(m, p)$, entonces

$$\begin{aligned} f(x_1, \dots, x_n, m, p) &= \prod_{i=1}^n f(x_i, m, p) = \prod_{i=1}^n \left[\binom{m}{x_i} p^{x_i} (1-p)^{m-x_i} \right] \\ &= \begin{cases} \left[\prod_{i=1}^n \binom{m}{x_i} \right] p^{\sum_{i=1}^n x_i} (1-p)^{nm-\sum_{i=1}^n x_i} & \text{si } x_i \in \{0, 1, \dots, m\}, \forall i = 1, 2, \dots, n \\ 0 & \text{fuera} \end{cases} \\ &= \left[\prod_{i=1}^n \binom{m}{x_i} \right] p^{\sum_{i=1}^n x_i} (1-p)^{nm-\sum_{i=1}^n x_i} h(x_1, \dots, x_n), \end{aligned}$$

siendo $h(x_1, \dots, x_n)$ la función indicador dada por $h(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, \dots, x_n \in \{0, 1, \dots, m\} \\ 0 & \text{fuera} \end{cases}$.

c) Si $X \rightsquigarrow P(\lambda)$, entonces

$$\begin{aligned} f(x_1, \dots, x_n, \lambda) &= \prod_{i=1}^n f(x_i, \lambda) = \left[\prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right] \\ &= \begin{cases} e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} & \text{si } x_i \in \mathbb{N} \cup \{0\}, \forall i = 1, 2, \dots, n \\ 0 & \text{fuera} \end{cases} \\ &= e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} h(x_1, \dots, x_n), \end{aligned}$$

siendo $h(x_1, \dots, x_n)$ la función indicador dada por $h(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, \dots, x_n \in \mathbb{N} \cup \{0\} \\ 0 & \text{fuera} \end{cases}$.

d) Si $X \rightsquigarrow U(a, b)$, entonces

$$\begin{aligned} f(x_1, \dots, x_n, a, b) &= \prod_{i=1}^n f(x_i, a, b) = \prod_{i=1}^n \left(\frac{1}{b-a} \right) \\ &= \begin{cases} \frac{1}{(b-a)^n} & \text{si } x_i \in (a, b), \forall i = 1, 2, \dots, n \\ 0 & \text{fuera} \end{cases} \\ &= \frac{1}{(b-a)^n} h_1(x_1, \dots, x_n, a) h_2(x_1, \dots, x_n, b), \end{aligned}$$

siendo $h_1(x_1, \dots, x_n)$ y $h_2(x_1, \dots, x_n, b)$ las funciones indicadores dadas por

$$h_1(x_1, \dots, x_n, a) = \begin{cases} 1 & \text{si } \min\{x_1, \dots, x_n\} > a \\ 0 & \text{en otro caso,} \end{cases}$$

y

$$h_2(x_1, \dots, x_n, b) = \begin{cases} 1 & \text{si } \max\{x_1, \dots, x_n\} < b \\ 0 & \text{en otro caso.} \end{cases}$$

e) Si $X \rightsquigarrow \exp(\lambda)$, entonces

$$\begin{aligned} f(x_1, \dots, x_n, \lambda) &= \prod_{i=1}^n f(x_i, \lambda) = \prod_{i=1}^n [\lambda e^{-\lambda x_i}] = \\ &= \begin{cases} \lambda^n e^{-\lambda \sum_{i=1}^n x_i} & \text{si } x_i \geq 0, \forall i = 1, 2, \dots, n \\ 0 & \text{fuera} \end{cases} \\ &= \lambda^n e^{-\lambda \sum_{i=1}^n x_i} h(x_1, \dots, x_n), \end{aligned}$$

siendo $h(x_1, \dots, x_n)$ la función indicador dada por $h(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, \dots, x_n \geq 0 \\ 0 & \text{fuera.} \end{cases}$

f) Si $X \rightsquigarrow \Gamma(a, p)$, entonces

$$\begin{aligned} f(x_1, \dots, x_n, a, p) &= \prod_{i=1}^n f(x_i, a, p) = \prod_{i=1}^n \left[\frac{p^a}{\Gamma(a)} e^{-p x_i} x_i^{a-1} \right] = \\ &= \begin{cases} \frac{p^{na}}{[\Gamma(a)]^n} e^{-p \sum_{i=1}^n x_i} \left(\prod_{i=1}^n x_i \right)^{a-1} & \text{si } x_i \geq 0, \forall i = 1, 2, \dots, n \\ 0 & \text{fuera} \end{cases} \\ &= \frac{p^{na}}{[\Gamma(a)]^n} e^{-p \sum_{i=1}^n x_i} \left(\prod_{i=1}^n x_i \right)^{a-1} h(x_1, \dots, x_n), \end{aligned}$$

siendo $h(x_1, \dots, x_n)$ la función indicador dada por $h(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, \dots, x_n \geq 0 \\ 0 & \text{fuera.} \end{cases}$

g) Si $X \rightsquigarrow N(\mu, \sigma)$, entonces

$$\begin{aligned} f(x_1, \dots, x_n, \mu, \sigma) &= \prod_{i=1}^n f(x_i, \mu, \sigma) = \prod_{i=1}^n \left(\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) \\ &= \frac{1}{(\sigma \sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}} \text{ si } x_i \in \mathbb{R}, \forall i = 1, 2, \dots, n. \end{aligned}$$

h) Si $X \rightsquigarrow B(a, b)$, entonces

$$f(x_1, \dots, x_n, a, b) = \prod_{i=1}^n f(x_i, a, b) = \prod_{i=1}^n \left(\frac{1}{\beta(a, b)} x_i^{a-1} (1 - x_i)^{b-1} \right)$$

$$= \begin{cases} (\frac{1}{\beta(a,b)})^n (\prod_{i=1}^n x_i)^{a-1} [\prod_{i=1}^n (1-x_i)]^{b-1} & \text{si } x_i \in (0,1), \forall i = 1, 2, \dots, n \\ 0 & \text{fuera} \end{cases}$$

$$= (\frac{1}{\beta(a,b)})^n (\prod_{i=1}^n x_i)^{a-1} [\prod_{i=1}^n (1-x_i)]^{b-1} h(x_1, \dots, x_n),$$

siendo $h(x_1, \dots, x_n)$ la función indicador dada por $h(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, \dots, x_n \in (0,1) \\ 0 & \text{fuera.} \end{cases}$

Bibliografía

1. E.L. Lehmann (1983). *Theory of Point Estimation*. John Wiley and Sons.
2. F.M. Dekking, C. Kraaikam, H.P. Lopuhaä and L.E. Meester (2005). *A Modern Introduction to Probability and Statistics. Understanding Why and How*. Springer.
3. M.H. DeGroot (1988). *Probability and Statistics*. Wilmington, Delaware, E.U.A. Addison-Wesley Iberoamericana, S.A.
4. V.K. Rohatgi and A.K. Md. Ehsanes Saleh (2001). *An Introduction to Probability and Statistics*. John Wiley and Sons.

Tema 8

Estimación puntual I. Introducción. Métodos para encontrar estimadores: método de los momentos, estimadores máximo-verosimiles, estimadores invariantes.

8.1 Estimación Puntual

Nos situamos en el modelo paramétrico, es decir, dispondremos de una muestra aleatoria simple X_1, X_2, \dots, X_n de una población X que se distribuye según $f(x, \theta)$, $\theta \in \Theta$, siendo θ un parámetro o vector de parámetros desconocidos. Θ es el espacio paramétrico, conjunto de valores posibles de θ . La estimación puntual persigue hacer un pronóstico de θ a partir de la muestra, utilizando algún estadístico $T = T(X_1, X_2, \dots, X_n)$, que en este contexto recibe el nombre de estimador.

Dado lo poco exigente que es la definición de estimador (estadístico) habrá estimadores con mejores y peores propiedades. La teoría de la estimación puntual se dirige a determinar criterios de estimación razonables y buscar para cada uno de ellos cuál sería el estimador óptimo, para los diferentes parámetros desconocidos de las distribuciones (o de modelos más complejos).

Ejemplo 8.1: Se desea estimar la media de las puntuaciones del curso 2020-2021, pero solo se dispone de 50 puntuaciones seleccionadas aleatoriamente. La media de la muestra (la estimación), es igual a 5.6 y atribuimos este valor a la media del curso completo.

Resumiendo,

Media poblacional curso 2020-2021, $\mu =$ Desconocida.

Estimador: Media muestral: \bar{X} .

Estimación de μ : 5.6.

Podemos utilizar como estimadores de la media de la población otros estadísticos de tendencia central como la moda o la mediana, pero NO todos los estimadores son apropiados. Los estimadores deben satisfacer ciertos requisitos, y por esta razón, como se ha dicho, interesa conocer sus propiedades a fin de utilizar los que sean adecuados según las circunstancias de la estimación.

8.1.1 Propiedades de los estimadores

A continuación se exponen tres propiedades que pueden cumplir (o no) los diferentes estimadores de un parámetro. Los siguientes conceptos serán extendidos en el capítulo próximo.

i) **Sesgo.** Se dice que un estimador es insesgado para estimar un cierto parámetro si la esperanza de la distribución del estimador es igual al parámetro.

Ya se vio anteriormente que la media muestral siempre es un estimador insesgado para la media poblacional y la cuasivarianza muestral siempre es un estimador insesgado para la varianza poblacional.

ii) **Consistencia.** Un estimador es consistente si aproxima (en probabilidad) el valor del parámetro cuanto mayor es n (tamaño de la muestra).

iii) **Eficiencia.** Diremos que un estimador es más eficiente que otro si la varianza de la distribución muestral del primero es menor que a la del segundo. Cuanto menor es la eficiencia, menor es la confianza en que el estadístico obtenido en la muestra aproxime al parámetro poblacional.

8.2 Estimadores invariantes.

Un estimador es invariante si se verifica que el estimador de una función del parámetro es igual a la función del estimador del parámetro. Es decir,

$$\widehat{f(\theta)} = f(\widehat{\theta}).$$

Por ejemplo si la varianza muestral es estimador de la varianza poblacional, cuando el método de estimación es invariante, la desviación típica muestral será estimador de la desviación típica poblacional. Existen estimadores invariantes a cambios de origen, cambios de escala, o cambios de origen y escala.

| Estimadores | C. origen | C. escala |
|-------------------------------------|---------------|---------------|
| \bar{x} (media muestral) | No invariante | No invariante |
| s^2 (varianza muestral) | Invariante | No invariante |
| s (desviación típica muestral) | Invariante | No invariante |
| ρ (Coeficiente de correlación) | Invariante | Invariante |

8.3 Estimación máximo verosímil

Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de una población $X \sim f(x, \theta)$, θ desconocido. Llamaremos estimador máximo verosímil de θ a $\hat{\theta}$ tal que:

$$f(x_1, x_2, \dots, x_n, \hat{\theta}) = \max_{\theta} f(x_1, x_2, \dots, x_n, \theta)$$

Entonces el estimador máximo verosímil es la función de la muestra que hacen más verosímiles los resultados observados. Puesto que nos adherimos a la máxima verosimilitud, los resultados observados debe suponerse que son los más probables, ya que bajo máxima verosimilitud sucede lo más probable. Para optimizar (maximizar) $f(x_1, x_2, \dots, x_n, \theta)$ en θ tenemos que anular la derivada (si existe) $\frac{\partial}{\partial \theta} f(x_1, x_2, \dots, x_n, \theta)$. Sin embargo, la función de verosimilitud es un producto de densidades, lo que la hace especialmente incómoda para la derivación. Pero como $\ln x$ es una función creciente, el valor de θ que maximiza $f(x_1, x_2, \dots, x_n, \theta)$ es el mismo que maximiza $\ln f(x_1, x_2, \dots, x_n, \theta)$. Por tanto, para determinar θ , resolveremos la llamada ecuación de verosimilitud.

$$\frac{\partial}{\partial \theta} \ln f(x_1, x_2, \dots, x_n, \theta) = 0.$$

No es necesario calcular la derivada segunda y ver que es negativa. Se puede demostrar que la raíz de la ecuación de verosimilitud (si existe) es un máximo. Si la función de verosimilitud no es derivable respecto de θ tendremos que acudir a otras técnicas que proporciona el cálculo para optimizar una función.

Ejemplo 8.2: Sea X_1, X_2, \dots, X_n muestra aleatoria simple de una población $X \sim B(1, p)$. Vamos a calcular el estimador máximo verosímil para p .

$$f(x_1, \dots, x_n, p) = \prod_{i=1}^n f(x_i, p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i},$$

$$\ln f(x_1, \dots, x_n, p) = \sum_{i=1}^n x_i \cdot \ln p + (n - \sum_{i=1}^n x_i) \ln(1-p),$$

$$\frac{\partial}{\partial \theta} \ln f(x_1, x_2, \dots, x_n, p) = \frac{\sum_{i=1}^n x_i}{p} + \frac{n - \sum_{i=1}^n x_i}{1-p} (-1) = \frac{\sum_{i=1}^n x_i - np}{p(1-p)} = 0$$

$$\Rightarrow \sum_{i=1}^n x_i - np = 0 \Rightarrow \hat{p} = \frac{\sum_{i=1}^n x_i}{n} = \bar{X}.$$

Ejemplo 8.3: Sea X_1, \dots, X_n muestra aleatoria simple de una población X . Determinar el estimador máximo verosímil para:

- a) p si $X \sim \text{Bin}(m, p)$.
- b) λ si $X \sim P(\lambda)$.
- c) θ si $X \sim U(0, \theta)$.
- d) a si $X \sim \Gamma(a, p)$, p conocido.

e) p si $X \sim \Gamma(a, p)$, a conocido.

f) μ si $X \sim N(\mu, \sigma_0)$.

g) σ si $X \sim N(\mu_0, \sigma)$.

h) $X \sim$ Cauchy centrada en θ .

i) θ si $X \sim B(\theta, 1)$.

a) p si $X \sim \text{Bin}(m, p)$.

Sea X_1, X_2, \dots, X_n muestra aleatoria simple de una población $X \sim \text{Bin}(m, p)$. Vamos a calcular el estimador máximo verosímil para p .

$$f(x_1, \dots, x_n, p) = \prod_{i=1}^n f(x_i, p) = \prod_{i=1}^n \binom{m}{x_i} p^{x_i} (1-p)^{m-x_i} = \left[\prod_{i=1}^n \binom{m}{x_i} \right] p^{\sum_{i=1}^n x_i} (1-p)^{nm - \sum_{i=1}^n x_i},$$

$$\ln f(x_1, \dots, x_n, p) = \sum_{i=1}^n \ln \binom{m}{x_i} + \sum_{i=1}^n x_i \cdot \ln p + (nm - \sum_{i=1}^n x_i) \ln(1-p),$$

$$\frac{\partial}{\partial p} \ln f(x_1, x_2, \dots, x_n, p) = \frac{\sum_{i=1}^n x_i}{p} + \frac{nm - \sum_{i=1}^n x_i}{1-p} (-1) = \frac{\sum_{i=1}^n x_i - np}{p(1-p)} = 0$$

$$\Rightarrow \sum_{i=1}^n x_i - np = 0 \Rightarrow \hat{p} = \frac{\sum_{i=1}^n X_i}{mn} = \frac{\bar{X}}{m}.$$

b) p si $X \sim P(\lambda)$.

Sea X_1, X_2, \dots, X_n muestra aleatoria simple de una población $X \sim P(\lambda)$. Vamos a calcular el estimador máximo verosímil para λ .

$$f(x_1, \dots, x_n, \lambda) = \prod_{i=1}^n f(x_i, \lambda) = \prod_{i=1}^n \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right) = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!},$$

$$\ln f(x_1, \dots, x_n, \lambda) = -n\lambda + \sum_{i=1}^n x_i \cdot \ln \lambda - \ln \left(\prod_{i=1}^n x_i! \right),$$

$$\frac{\partial}{\partial \lambda} \ln f(x_1, x_2, \dots, x_n, \lambda) = -n + \frac{\sum_{i=1}^n x_i}{\lambda} = 0.$$

$$\Rightarrow \hat{\lambda} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}.$$

c) θ si $X \sim U(0, \theta)$.

Sea X_1, X_2, \dots, X_n muestra aleatoria simple de una población $X \sim U(0, \theta)$. Vamos a calcular el estimador máximo verosímil para θ .

$$f(x_1, x_2, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n \frac{1}{\theta} = \begin{cases} \frac{1}{\theta^n} & \text{si } x_1, x_2, \dots, x_n \in [0, \theta] \\ 0 & \text{fuera.} \end{cases}$$

$$= \begin{cases} \frac{1}{\theta^n} & \text{si } \min\{x_1, x_2, \dots, x_n\} \geq 0 \text{ y } \max\{x_1, x_2, \dots, x_n\} \leq \theta \\ 0 & \text{fuera.} \end{cases}$$

La función no es derivable respecto de θ . Para optimizar θ observamos que $\frac{1}{\theta^n}$ es decreciente de θ . Como queremos maximizar en θ , debemos asignarle el valor más pequeño posible. Como $\theta \geq \max\{x_1, x_2, \dots, x_n\}$, entonces $\hat{\theta} = \max\{X_1, X_2, \dots, X_n\}$.

Este ejemplo, ligeramente matizado sirve para probar que el estimador máximo verosímil puede no existir en un modelo dado.

Si modificamos la densidad uniforme escribiendo

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} & \text{si } x \in [0, \theta) \\ 0 & \text{fuera,} \end{cases}$$

un análisis como el previo nos lleva a que

$$f(x_1, x_2, \dots, x_n, \theta) = \begin{cases} \frac{1}{\theta^n} & \text{si } \min\{x_1, \dots, x_n\} \geq 0 \text{ y } \max\{x_1, \dots, x_n\} < \theta \\ 0 & \text{fuera.} \end{cases}$$

Entonces el estimador máximo verosímil es el menor valor posible para θ bajo la restricción $\theta > \max\{x_1, \dots, x_n\}$, de donde se concluye que no existe en este caso.

d) a si $X \sim \Gamma(a, p)$, p conocido.

Sea X_1, X_2, \dots, X_n muestra aleatoria simple de una población $X \sim \Gamma(a, p)$. Vamos a calcular el estimador máximo verosímil para a .

$$f(x_1, \dots, x_n, a) = \prod_{i=1}^n f(x_i, a) = \prod_{i=1}^n \left(\frac{p^a}{\Gamma(a)} e^{-px_i} x_i^{a-1} \right) = \frac{p^{na}}{[\Gamma(a)]^n} e^{-p \sum_{i=1}^n x_i} \left(\prod_{i=1}^n x_i \right)^{a-1},$$

$$\ln f(x_1, \dots, x_n, a) = na \ln p - n \ln \Gamma(a) - p \sum_{i=1}^n x_i + (a-1) \sum_{i=1}^n \ln x_i,$$

$$\frac{\partial}{\partial a} \ln f(x_1, x_2, \dots, x_n, a) = n \ln p - n \frac{\Gamma'(a)}{\Gamma(a)} + \sum_{i=1}^n \ln x_i = 0,$$

es decir,

$$\frac{\Gamma'(a)}{\Gamma(a)} = \frac{1}{n} \sum_{i=1}^n \ln x_i$$

No hay expresión funcional para \hat{a} , pero conocida las observaciones se puede resolver numéricamente la ecuación.

En diversas colecciones publicadas de tablas matemáticas, se incluyen tablas de la función $\frac{\Gamma'(a)}{\Gamma(a)}$, que se denomina función digamma. Para cualesquiera valores concretos x_1, \dots, x_n el único valor de a que satisface la ecuación anterior se debe determinar consultando esas tablas o realizando un análisis numérico de la función digamma. Este valor será el estimador máximo verosímil de a .

e) p si $X \sim \Gamma(a, p)$, a conocido.

Sea X_1, X_2, \dots, X_n muestra aleatoria simple de una población $X \sim \Gamma(a, p)$. Vamos a calcular el estimador máximo verosímil para p .

$$f(x_1, \dots, x_n, p) = \prod_{i=1}^n f(x_i, p) = \prod_{i=1}^n \left(\frac{p^a}{\Gamma(a)} e^{-px_i} x_i^{a-1} \right) = \frac{p^{na}}{[\Gamma(a)]^n} e^{-p \sum_{i=1}^n x_i} \left(\prod_{i=1}^n x_i \right)^{a-1},$$

$$\ln f(x_1, \dots, x_n, p) = na \ln p - n \ln(\Gamma(a)) - p \sum_{i=1}^n x_i + (a-1) \sum_{i=1}^n \ln x_i,$$

$$\frac{\partial}{\partial p} \ln f(x_1, x_2, \dots, x_n, p) = \frac{na}{p} - \sum_{i=1}^n x_i = 0 \Rightarrow \hat{p} = \frac{na}{\sum_{i=1}^n x_i} = \frac{a}{\bar{X}}.$$

f) μ si $X \sim N(\mu, \sigma)$ σ conocido.

Sea X_1, X_2, \dots, X_n muestra aleatoria simple de una población $X \sim N(\mu, \sigma)$. Vamos a calcular el estimador máximo verosímil para μ .

$$f(x_1, \dots, x_n, \mu) = \prod_{i=1}^n f(x_i, \mu) = \prod_{i=1}^n \left(\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) = \frac{1}{(\sigma \sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}},$$

$$\ln f(x_1, \dots, x_n, \mu) = -n \ln \sigma - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2,$$

$$\frac{\partial}{\partial \mu} \ln f(x_1, x_2, \dots, x_n, \mu) = \frac{-2}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu)(-1) = 0 \Rightarrow \sum_{i=1}^n x_i - n\mu = 0 \Rightarrow \hat{\mu} = \bar{X}.$$

g) σ si $X \sim N(\mu, \sigma)$, μ conocido.

Sea X_1, X_2, \dots, X_n muestra aleatoria simple de una población $X \sim N(\mu, \sigma)$. Vamos a calcular el estimador máximo verosímil para σ .

$$f(x_1, \dots, x_n, \sigma) = \prod_{i=1}^n f(x_i, \sigma) = \prod_{i=1}^n \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}},$$

$$\ln f(x_1, \dots, x_n, \sigma) = -n \ln \sigma - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2,$$

$$\frac{\partial}{\partial \sigma} \ln f(x_1, x_2, \dots, x_n, \sigma) = \frac{-n}{\sigma} - \frac{1}{2} (-2\sigma^{-3}) \sum_{i=1}^n (x_i - \mu)^2 = 0$$

$$\Rightarrow \frac{-n\sigma^2 + \sum_{i=1}^n (x_i - \mu)^2}{\sigma^3} = 0 \Rightarrow \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n},$$

y por tanto

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{n}}.$$

h) $X \sim$ Cauchy centrada en θ .

Supongamos que las variables X_1, \dots, X_n constituyen una muestra aleatoria simple de una distribución Cauchy centrada en el punto desconocido θ ($-\infty < \theta < \infty$). La función de densidad es la siguiente:

$$f(x, \theta) = \frac{1}{\pi[1 + (x - \theta)^2]}, \text{ para } -\infty < x < \infty.$$

Obtendremos el estimador máximo verosímil de θ .

La función de verosimilitud es:

$$f(x_1, \dots, x_n, \theta) = \frac{1}{\pi^n \prod_{i=1}^n [1 + (x_i - \theta)^2]}.$$

Por tanto, el estimador máximo verosímil de θ será el valor que minimice

$$\prod_{i=1}^n [1 + (x_i - \theta)^2] \tag{8.1}$$

Para cualesquiera valores de x_1, \dots, x_n el valor de θ que minimiza la expresión 8.1 se debe determinar por medio de cálculos numéricos.

i) si $X \sim B(\theta, 1)$.

Sea X_1, X_2, \dots, X_n muestra aleatoria simple de una población $X \sim B(\theta, 1)$ con función de densidad

$$f(x, \theta) = \begin{cases} \theta \cdot x^{\theta-1} & \text{si } x \in (0, 1) \\ 0 & \text{fuera.} \end{cases}$$

Entonces

$$f(x_1, \dots, x_n, \theta) = \begin{cases} \theta^n (\prod_{i=1}^n x_i)^{\theta-1} & \text{si } x_1, \dots, x_n \in (0, 1) \\ 0 & \text{fuera.} \end{cases}$$

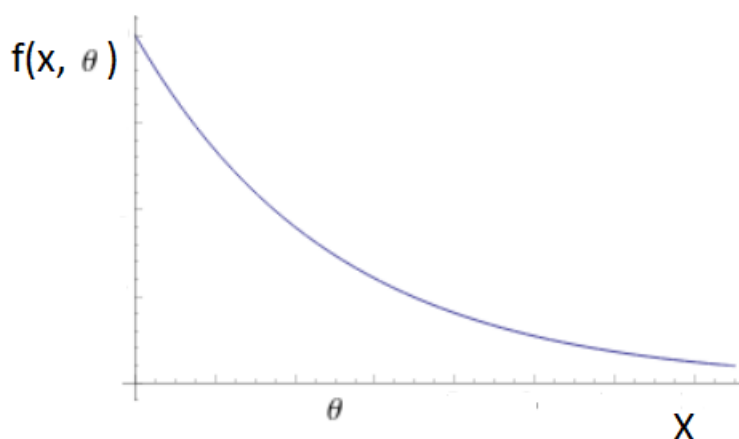
$$\ln f(x_1, \dots, x_n, \theta) = n \ln \theta + (\theta - 1) \ln \left(\prod_{i=1}^n x_i \right) = n \ln \theta + (\theta - 1) \sum_{i=1}^n \ln x_i.$$

$$\frac{\partial}{\partial \theta} \ln f(x_1, \dots, x_n, \theta) = \frac{n}{\theta} + \sum_{i=1}^n \ln x_i = 0 \Rightarrow \hat{\theta} = \frac{-n}{\sum_{i=1}^n \ln X_i},$$

es el estimador máximo verosímil.

Ejemplo 8.4: Sea X_1, \dots, X_n muestra aleatoria simple de $X \sim f(x, \theta)$ siendo:

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \text{si } x \geq 0 \\ 0 & \text{fuera,} \end{cases}$$



Calcular el estimador máximo verosímil de θ .

$$f(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n \left(\frac{1}{\theta} e^{-\frac{x_i}{\theta}} \right) = \frac{1}{\theta^n} e^{-\frac{\sum_{i=1}^n x_i}{\theta}},$$

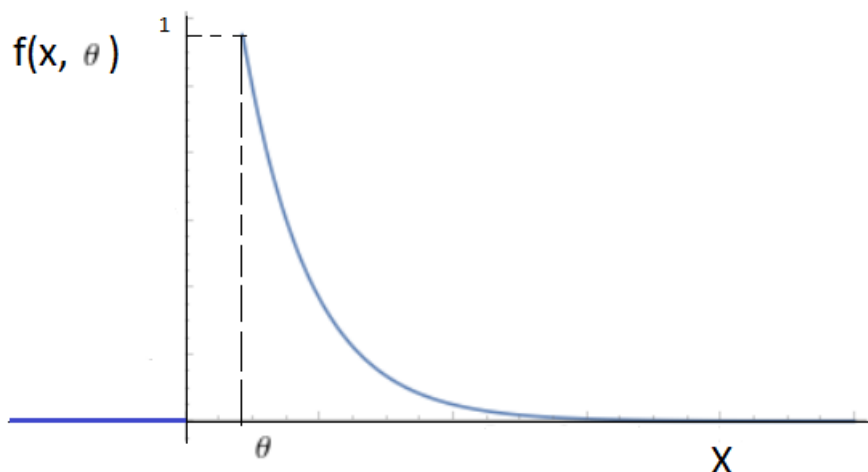
$$\ln f(x_1, \dots, x_n, \theta) = -n \ln \theta - \frac{\sum_{i=1}^n x_i}{\theta},$$

$$\frac{\partial}{\partial \theta} \ln f(x_1, x_2, \dots, x_n, \theta) = \frac{-n\theta + \sum_{i=1}^n x_i}{\theta^2} = 0$$

$$\Rightarrow -n\theta + \sum_{i=1}^n x_i = 0 \Rightarrow \hat{\theta} = \bar{X}.$$

Ejemplo 8.5: Sea X_1, \dots, X_n muestra aleatoria simple de $X \sim f(x, \theta)$ siendo:

$$f(x, \theta) = \begin{cases} e^{-x} e^{\theta} & \text{si } x \geq \theta \\ 0 & \text{fuera.} \end{cases}$$



Calcular el estimador máximo verosímil de θ .

$$\begin{aligned} f(x_1, \dots, x_n, \theta) &= \begin{cases} e^{-\sum_{i=1}^n x_i} e^{n\theta} & \text{si } x_1, x_2, \dots, x_n \geq \theta \\ 0 & \text{fuera.} \end{cases} \\ &= \begin{cases} e^{-\sum_{i=1}^n x_i} e^{n\theta} & \text{si } \min\{x_1, x_2, \dots, x_n\} \geq \theta \\ 0 & \text{fuera.} \end{cases} \end{aligned}$$

No es derivable respecto de θ . $f(x_1, \dots, x_n, \theta)$ es una función creciente de θ ya que $e^{n\theta}$ lo es. Para maximizarla debemos asignar a θ el valor más grande posible. Dado que $\theta \leq \min\{x_1, \dots, x_n\}$. El estimador máximo verosímil es $\hat{\theta} = \min\{X_1, \dots, X_n\}$.

Ejemplo 8.6: Sea X_1, \dots, X_n muestra aleatoria simple de $X \sim f(x, \theta)$ siendo:

$$f(x, \theta) = ke^{-\frac{|x|}{\theta}} \text{ si } x \in \mathbb{R}, \theta \geq 0.$$

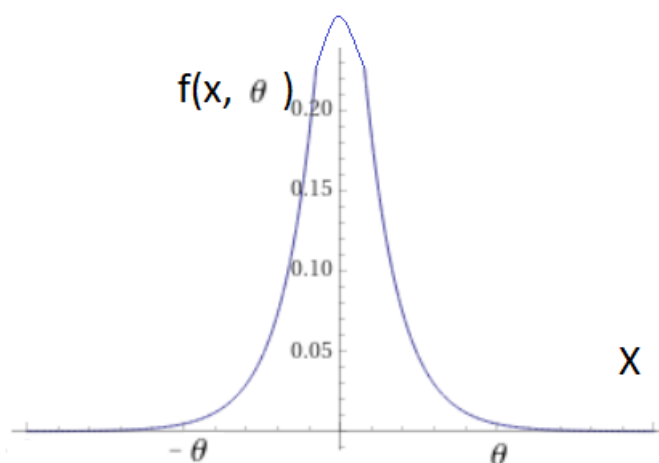
Vamos a calcular el estimador máximo verosímil para θ .

$$f(x, \theta) = ke^{-\frac{|x|}{\theta}} \text{ si } x \in \mathbb{R}, \theta \geq 0.$$

$$= \begin{cases} ke^{-\frac{x}{\theta}} & \text{si } x \geq 0, \theta \geq 0 \\ ke^{\frac{x}{\theta}} & \text{si } x < 0, \theta \geq 0. \end{cases}$$

En primer lugar, vamos a calcular el valor de k .

$$1 = \int_{-\infty}^0 ke^{\frac{x}{\theta}} dx + \int_0^{\infty} ke^{-\frac{x}{\theta}} dx \Rightarrow k = \frac{1}{2\theta}.$$



Una vez conocemos al completo la función de densidad, procedemos a calcular el estimador máximo verosímil.

$$f(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n \left(\frac{1}{2\theta} e^{-\frac{|x_i|}{\theta}} \right) = \frac{1}{(2\theta)^n} e^{-\frac{\sum_{i=1}^n |x_i|}{\theta}},$$

$$\ln f(x_1, \dots, x_n, \theta) = -n \ln 2 - n \ln \theta - \frac{1}{\theta} \sum_{i=1}^n |x_i|,$$

$$\frac{\partial}{\partial \theta} \ln f(x_1, x_2, \dots, x_n, \theta) = \frac{-n}{\theta} + \frac{\sum_{i=1}^n |x_i|}{\theta^2} = 0$$

$$\Rightarrow -n\theta + \sum_{i=1}^n |x_i| = 0 \Rightarrow \hat{\theta} = \frac{\sum_{i=1}^n |X_i|}{n} = \overline{|X|}.$$

8.3.1 Invarianza del estimador máximo verosímil.

Supóngase que las variables X_1, \dots, X_n constituyen una muestra aleatoria simple de una distribución cuya función de probabilidad o función de densidad es $f(x_1, \dots, x_n, \theta)$ con parámetro θ desconocido y sea $\hat{\theta}$ el estimador máximo verosímil de θ . Entonces, para cualesquiera valores x_1, \dots, x_n la función de verosimilitud $f(x_1, \dots, x_n, \theta)$ se maximiza cuando $\theta = \hat{\theta}$.

Supóngase ahora que se cambia el parámetro de la distribución como sigue: en lugar de expresar la función de probabilidad o la función de densidad $f(x, \theta)$ en función del parámetro θ , se expresa en función de un nuevo parámetro $t = g(\theta)$, donde g es una función biunívoca de θ . Se define $\theta = h(t)$ como la correspondiente función inversa. Entonces, la función de verosimilitud será $f(x_1, \dots, x_n, h(t))$.

El estimador máximo verosímil $\widehat{g(\theta)}$ será igual al valor de t que maximice $f(x_1, \dots, x_n, h(t))$. Puesto que $f(x_1, \dots, x_n, \theta)$ se maximiza cuando $\theta = \hat{\theta}$, resulta que $f(x_1, \dots, x_n, h(t))$ se maximiza cuando $h(t) = \hat{\theta}$. Por tanto, el estimador máximo verosímil \hat{t} debe satisfacer la relación $\widehat{h(t)} = \hat{\theta}$ o equivalentemente, $\hat{t} = \widehat{g(\theta)}$. Se ha establecido por tanto, la siguiente propiedad, que se denomina propiedad de invarianza de los estimadores máximo verosímiles.

Proposición 8.1: Sea X_1, X_2, \dots, X_n muestra aleatoria simple de $X \sim f(x, \theta)$. Si $\hat{\theta}$ es el estimador máximo verosímil de θ y $g(\theta)$ es una función biunívoca de θ , entonces

$$\widehat{g(\theta)} = g(\hat{\theta}).$$

Es decir, el estimador máximo verosímil de $g(\theta)$ es g del estimador máximo verosímil de θ .

Ejemplo 8.7: Sea X_1, X_2, \dots, X_n m.a.s. de $X \sim N(\mu, \sigma)$, $\widehat{\sigma^2} = S^2$ y $\widehat{\sigma} = S$ ya que $\sigma = \sqrt{\sigma^2}$ es función biunívoca de σ^2 .

Ejemplo 8.8: Sea X_1, X_2, \dots, X_n m.a.s. de $X \sim P(\lambda)$. Obtener el estimador máximo verosímil para $P(X = 0)$.

$g(\lambda) = P(X = 0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda}$ que es función biyectiva de λ . Luego por la invarianza de estimador máximo verosímil, $\widehat{g(\lambda)} = \widehat{(e^{-\lambda})} = e^{-\hat{\lambda}} = e^{-\bar{x}}$

Por tanto, si las observaciones en una realización muestral de tamaño $n = 8$, fueran 1, 3, 5, 5, 2, 8, 6, 4, la estimación de λ sería,

$$\lambda^* = \bar{X} = \frac{34}{8} = 4,25,$$

y la de $g(\lambda)$

$$P(X = 0)^* = e^{-4,25}.$$

Si no tuvieramos información de que la distribución fuera de Poisson, $P(X = 0)^* = \frac{0}{8} = 0$, $P(X = 5)^* = \frac{2}{8} = 0,25$. Con información sobre la distribución (Poisson), $P(X = 5)^* = e^{-4,25} \frac{4,25^5}{5!}$.

Ejemplo 8.9: Sea 2, 5, 3, 1, 0, 7, 1, 2, 0, 8 una realización muestral de $X \sim \exp(\lambda)$, obtener la estimación máximo verosímil de $P(X \geq 1)$. Utilizamos el principio de invarianza.

$$P(X \geq 1) = \int_1^{\infty} \lambda e^{-\lambda x} dx = -e^{-\lambda x} \Big|_1^{\infty} = e^{-\lambda}.$$

Luego el estimador máximo verosímil de $P(X \geq 1)$ es:

$$\widehat{P(X \geq 1)} = e^{-\hat{\lambda}} = e^{-\frac{1}{\bar{x}}},$$

y la estimación para la realización muestral obtenida es:

$$P(X \geq 1)^* = e^{-\frac{1}{2.9}}.$$

8.3.2 Estimación máximo verosímil en el caso de varios parámetros desconocidos

Sea X_1, X_2, \dots, X_n muestra aleatoria simple de una población $X \sim f(x, \theta_1, \dots, \theta_k)$, siendo $\theta_1, \dots, \theta_k$ parámetros desconocidos. Para obtener los estimadores máximo verosímiles de $\theta_1, \dots, \theta_k$ en el caso de que $f(x_1, \dots, x_n, \theta_1, \dots, \theta_k)$ sea función derivable respecto de cada uno de los parámetros, anularemos el gradiente de $\ln f(x_1, \dots, x_n, \theta_1, \dots, \theta_k)$. Es decir, los estimadores máximo verosímiles se obtendrán como raíces del sistema de ecuaciones de verosimilitud.

$$\frac{\partial}{\partial \theta_1} \ln f(x_1, \dots, x_n, \theta_1, \dots, \theta_k) = 0$$

$$\frac{\partial}{\partial \theta_2} \ln f(x_1, \dots, x_n, \theta_1, \dots, \theta_k) = 0$$

...

$$\frac{\partial}{\partial \theta_k} \ln f(x_1, \dots, x_n, \theta_1, \dots, \theta_k) = 0.$$

Puede demostrarse que la solución de este sistema hace definida negativa la matriz Hessiana en ese punto. Por tanto, en las aplicaciones no es necesario calcular la matriz Hessiana. Si la función de verosimilitud no es derivable de algún parámetro tenemos que analizar el crecimiento-decrecimiento en dicho parámetro.

Ejemplo 8.10: Sea X_1, X_2, \dots, X_n muestra aleatoria simple de una población $X \sim N(\mu, \sigma)$ con μ, σ desconocidos. Obtener el estimador máximo verosímil de μ y de σ .

$$f(x_1, \dots, x_n, \mu, \sigma) = \prod_{i=1}^n f(x_i, \mu, \sigma) = \prod_{i=1}^n \left(\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) = \frac{1}{(\sigma \sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}},$$

$$\ln f(x_1, \dots, x_n, \mu, \sigma) = -n \cdot \ln(\sqrt{2\pi}) - n \cdot \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2,$$

$$\frac{\partial}{\partial \mu} \ln f(x_1, x_2, \dots, x_n, \mu, \sigma) = \frac{-2}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)(-1) = 0 \Rightarrow \sum_{i=1}^n x_i - n\mu = 0 \Rightarrow \hat{\mu} = \bar{X}$$

$$\begin{aligned} \frac{\partial}{\partial \sigma} \ln f(x_1, x_2, \dots, x_n, \mu, \sigma) &= \frac{-n}{\sigma} + \frac{2}{2\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 = 0 \Rightarrow -n\sigma^2 + \sum_{i=1}^n (x_i - \mu)^2 = 0 \\ &\Rightarrow \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = S^2 \text{ ó } \hat{\sigma} = S. \end{aligned}$$

Comprobemos que efectivamente son máximos demostrando que la matriz Hessiana en ellos es definida negativa.

$$\begin{aligned} H[\ln f(X_1, X_2, \dots, X_n, \mu, \sigma)] &= \\ \begin{pmatrix} \frac{\partial^2}{\partial \mu^2} \ln f(x_1, x_2, \dots, x_n, \mu, \sigma) & \frac{\partial^2}{\partial \mu \partial \sigma} \ln f(x_1, x_2, \dots, x_n, \mu, \sigma) \\ \frac{\partial^2}{\partial \sigma \partial \mu} \ln f(x_1, x_2, \dots, x_n, \mu, \sigma) & \frac{\partial^2}{\partial \sigma^2} \ln f(x_1, x_2, \dots, x_n, \mu, \sigma) \end{pmatrix} &= \\ \begin{pmatrix} \frac{-n}{\sigma^2} & \frac{\sum_{i=1}^n x_i - n\mu}{\sigma^2}(-2) \\ \frac{\sum_{i=1}^n x_i - n\mu}{\sigma^2}(-2) & \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^n (x_i - n\mu)^2 \end{pmatrix}. \end{aligned}$$

Por tanto,

$$H[\ln f(X_1, X_2, \dots, X_n, \hat{\mu}, \hat{\sigma})] =$$

$$\begin{aligned} \begin{pmatrix} \frac{-n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{\hat{\sigma}^2} - \frac{3ns^2}{\hat{\sigma}^4} \end{pmatrix} &= \\ \begin{pmatrix} \frac{-n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{\hat{\sigma}^2} - \frac{2n}{\hat{\sigma}^2} \end{pmatrix} &= \begin{pmatrix} \frac{-n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{-n}{\hat{\sigma}^2} \end{pmatrix}. \end{aligned}$$

Para que una matriz $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ sea definida negativa debe ocurrir que $a < 0$ y que $\begin{vmatrix} a & b \\ c & d \end{vmatrix} > 0$. En efecto, $\frac{-n}{\hat{\sigma}^2} < 0$ y

$$\begin{vmatrix} \frac{-n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{\hat{\sigma}^2} - \frac{2n}{\hat{\sigma}^2} \end{vmatrix} = \frac{2n^2}{s^4} > 0.$$

Luego $\hat{\mu} = \bar{X}$ y $\hat{\sigma}^2 = S^2$ son máximos y son el estimadores máximos verosímiles.

La propiedad de invarianza vista anteriormente, se puede extender a funciones de un vector de parámetros θ . Supóngase que $\theta = (\theta_1, \dots, \theta_k)$ es un vector de k parámetros reales. Si $t = g(\theta_1, \dots, \theta_k)$ es una función real de $\theta_1, \dots, \theta_k$, entonces t se puede considerar como una componente de una transformación biunívoca del conjunto de parámetros $\theta_1, \dots, \theta_k$ a un nuevo conjunto de k parámetros reales. Por tanto, si $\hat{\theta}_1, \dots, \hat{\theta}_k$ son los estimadores máximo verosímiles de $\theta_1, \dots, \theta_k$, de la propiedad de invarianza resulta que el estimador máximo verosímil de t es $\hat{t} = g(\hat{\theta}_1, \dots, \hat{\theta}_k)$.

8.4 Consistencia de un estimador

La consistencia de un estimador tiene que ver con su comportamiento cuando se incrementa el tamaño muestral.

Definición 8.1: Dados X_1, X_2, \dots, X_n una muestra aleatoria simple de una población X que se distribuye según $f(x, \theta)$, θ desconocido, diremos que el estadístico $T = T(X_1, X_2, \dots, X_n)$ es consistente para estimar θ si converge en probabilidad hacia θ a medida que n (tamaño muestral) tiende a infinito, es decir, si

$$\lim_{n \rightarrow \infty} P(|T(X_1, \dots, X_n) - \theta| > \varepsilon) = 0, \forall \varepsilon > 0.$$

La convergencia en probabilidad de $T(X_1, \dots, X_n)$ a θ se nota

$$T(X_1, \dots, X_n) \xrightarrow{P} \theta.$$

Al utilizar un estimador consistente el pronóstico que se haga de θ estará más y más próximo a θ a medida de que el tamaño muestral crezca (la calidad de la estimación mejora con el aumento de la información).

Algunos ejemplos: Como consecuencia de la ley débil de los grandes números si $\theta = EX$, $T = \bar{X}$ es un estimador consistente para θ pues tal ley afirma que $\bar{X} \xrightarrow{P} EX = \theta$.

Por tanto si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $B(1, p)$ entonces \bar{X} (frecuencia relativa) es consistente para estimar p .

Ejemplo 8.11: Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $P(\lambda)$ entonces \bar{X} es consistente para estimar λ .

Ejemplo 8.12: Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $N(\mu, \sigma)$ entonces \bar{X} es consistente para estimar μ , S^2 y \hat{S}^2 son consistentes para estimar σ^2 , S y \hat{S} son consistentes para estimar σ .

Ejemplo 8.13: Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $U(a, b)$ entonces $\min\{X_1, X_2, \dots, X_n\}$ es consistente para estimar a y $\max\{X_1, X_2, \dots, X_n\}$ es consistente para estimar b .

Ejemplo 8.14: Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $\exp(\lambda)$ entonces $\bar{X} \xrightarrow{P} EX = \frac{1}{\lambda}$, luego $\frac{1}{\bar{X}} \xrightarrow{P} \frac{1}{\frac{1}{\lambda}} = \lambda$ y entonces $\frac{1}{\bar{X}}$ es consistente para estimar λ .

8.5 Estimación por el método de los momentos.

Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de $X \sim f(x, \theta_1, \theta_2)$ siendo θ_1 y θ_2 dos parámetros desconocidos. El método de los momentos es un método de sencilla aplicación, pero solo garantiza que los estimadores obtenidos por dicho método son consistentes. Para aplicarlo escribimos los parámetros θ_1 y θ_2 como funciones de $E(X)$ y $Var(X)$:

$$\theta_1 = g_1(E(X), Var(X)),$$

y

$$\theta_2 = g_2(E(X), Var(X)).$$

Entonces los estimadores por el método de los momentos se obtienen sustituyendo en g_1 y g_2 , $E(X)$ por \bar{X} y $Var(X)$ por S^2 ,

$$\hat{\theta}_1 = g_1(\bar{X}, S^2),$$

y

$$\hat{\theta}_2 = g_2(\bar{X}, S^2).$$

Como

$$\bar{X} \xrightarrow{P} E(X),$$

y

$$S^2 \xrightarrow{P} Var(X),$$

por la ley débil de los grandes números, si g_1 y g_2 son funciones continuas de $E(X)$ y $Var(X)$, entonces:

$$\hat{\theta}_1 = g_1(\bar{X}, S^2) \xrightarrow{P} g_1(E(X), Var(X)) = \theta_1,$$

y

$$\hat{\theta}_2 = g_2(\bar{X}, S^2) \xrightarrow{P} g_2(E(X), Var(X)) = \theta_2.$$

Así que se garantiza la consistencia de los estimadores obtenidos por el método de los momentos.

Notese que este es un método de obtención de estimadores y que bajo ningún concepto presupone que $EX = \bar{X}$ o que $VarX = S^2$ pues como se insistió, parámetros poblacionales y estimadores no deben confundirse. Precisamente el hecho de que si n es grande \bar{X} se aproxime a EX y S^2 se aproxime a $VarX$ esto le da sentido al método.

Ejemplo 8.15: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de

$$X \sim B(1, p).$$

Para calcular un estimador de p por el método de los momentos, escribimos p en función de $E(X)$,

$$p = E(X),$$

y entonces el estimador por el método de los momentos es

$$\hat{p} = \bar{X}.$$

Es consistente y además es estimador máximo verosímil.

Ejemplo 8.16: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de

$$X \sim \text{Bin}(m, p).$$

Para calcular un estimador de p por el método de los momentos, escribimos p en función de $E(X)$,

$$p = \frac{E(X)}{m},$$

y entonces el estimador por el método de los momentos es

$$\hat{p} = \frac{\bar{X}}{m}.$$

Es consistente y además es estimador máximo verosímil.

Ejemplo 8.17: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de

$$X \sim P(\lambda).$$

Para calcular un estimador de λ por el método de los momentos, escribimos λ en función de $E(X)$,

$$\lambda = E(X),$$

y entonces el estimador por el método de los momentos es

$$\hat{\lambda} = \bar{X}.$$

Es consistente y además es estimador máximo verosímil.

Ejemplo 8.18: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de

$$X \sim U(0, \theta).$$

Para calcular un estimador de θ por el método de los momentos, escribimos θ en función de $E(X)$,

$$\theta = 2E(X), \text{ ya que } E(X) = \frac{\theta}{2}.$$

Entonces el estimador por el método de los momentos es

$$\hat{\theta} = 2\bar{X}.$$

Es consistente pero no es estimador máximo verosímil, ya que éste es $\hat{\theta} = \max\{X_1, X_2, \dots, X_n\}$.

Ejemplo 8.19: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de

$$X \sim \exp(\lambda).$$

Para calcular un estimador de λ por el método de los momentos, escribimos λ en función de $E(X)$,

$$\lambda = \frac{1}{E(X)},$$

y entonces el estimador por el método de los momentos es

$$\hat{\lambda} = \frac{1}{\bar{X}}.$$

Es consistente y además es estimador máximo verosímil.

Ejemplo 8.20: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de

$$X \sim N(\mu, \sigma),$$

μ y σ ambos desconocidos. Para calcular un estimador de μ y otro de σ por el método de los momentos, escribimos μ y σ en función de $E(X)$ y $Var(X)$,

$$\mu = E(X),$$

y

$$\sigma^2 = Var(X), \text{ por lo que } \sigma = \sqrt{Var(X)}.$$

Entonces el estimador por el método de los momentos es

$$\hat{\mu} = \bar{X},$$

$$\widehat{\sigma^2} = S^2,$$

y

$$\hat{\sigma} = S.$$

$\hat{\mu}$ es consistente y también estimador máximo verosímil; S^2 es consistente para σ^2 y es estimador máximo verosímil y S es consistente para σ y también es estimador máximo verosímil.

Ejemplo 8.21: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de

$$X \sim \Gamma(a, p),$$

a y p ambos desconocidos. Para calcular un estimador de a y otro de p por el método de los momentos, escribimos a y p en función de $E(X)$ y $Var(X)$. Entonces, dado que

$$E(X) = \frac{a}{p},$$

y

$$Var(X) = \frac{a}{p^2},$$

se tiene

$$\frac{E(X)}{\text{Var}(X)} = p,$$

y

$$a = pE(X) = \frac{E^2(X)}{\text{Var}(X)}.$$

Entonces, los estimadores por el método de los momentos son:

$$\hat{a} = \frac{\bar{X}^2}{S^2},$$

y

$$\hat{p} = \frac{\bar{X}}{S^2}.$$

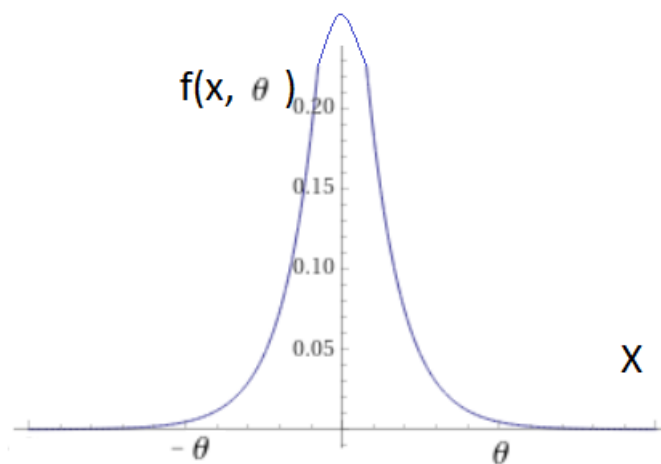
Ambos estimadores son necesariamente consistentes pero no son estimadores máximos verosímiles.

Ejemplo 8.22: Sea X_1, \dots, X_n muestra aleatoria simple de $X \sim f(x, \theta)$ siendo:

$$f(x, \theta) = ke^{-\frac{|x|}{\theta}} \text{ si } x \in \mathbb{R}, \theta \geq 0.$$

Vamos a calcular el estimador por el método de los momentos para θ . Primero, vamos a calcular el valor de k .

$$\begin{aligned} f(x, \theta) &= ke^{-\frac{|x|}{\theta}} \text{ si } x \in \mathbb{R}, \theta \geq 0 \\ &= \begin{cases} ke^{-\frac{x}{\theta}} & \text{si } x > 0, \theta \geq 0 \\ ke^{\frac{x}{\theta}} & \text{si } x < 0, \theta \geq 0 \end{cases} \\ 1 &= \int_{-\infty}^0 ke^{\frac{x}{\theta}} dx + \int_0^{\infty} ke^{-\frac{x}{\theta}} dx \Rightarrow k = \frac{1}{2\theta} \end{aligned}$$



Una vez conocemos al completo la función de densidad, procedemos a calcular el estimador por el método de los momentos.

$$\begin{aligned}
 E(|X_i|) &= \int_{-\infty}^{\infty} |x|f(x, \theta)dx = \underbrace{\int_{-\infty}^0 -xke^{\frac{x}{\theta}}dx}_{t=-x; dt=-dx} + \int_0^{\infty} xke^{\frac{-x}{\theta}}dx \\
 &= \int_{\infty}^0 tke^{\frac{-t}{\theta}}(-dt) + \int_0^{\infty} xke^{\frac{-x}{\theta}}dx \\
 &= \int_0^{\infty} tke^{\frac{-t}{\theta}}dt + \int_0^{\infty} xke^{\frac{-x}{\theta}}dx \\
 &= 2 \int_0^{\infty} xke^{\frac{-x}{\theta}}dx \\
 &= 2k \int_0^{\infty} xe^{\frac{-x}{\theta}}dx
 \end{aligned}$$

Recordamos que,

$$\int_0^{\infty} x^{a-1}e^{-px}dx = \frac{(a-1)!}{p},$$

Aplicándolo a nuestro ejercicio con $a = 2$ y $p = \frac{1}{\theta}$, tenemos,

$$\begin{aligned}
 E(|X_i|) &= 2k \int_0^{\infty} xe^{\frac{-x}{\theta}}dx \\
 &= 2k \frac{(2-1)!}{(\frac{1}{\theta})^2} \\
 &= \frac{2k}{\frac{1}{\theta^2}} = \theta.
 \end{aligned}$$

Dado que $E(|X_i|) = \theta$, entonces, el estimador por el método de los momentos para θ es $\hat{\theta} = \overline{|X|} = \frac{\sum_{i=1}^n |X_i|}{n}$. Si por ejemplo, estuviéramos bajo la siguiente realización muestral $(-1, 2, -2, 3, 5)$, la estimación por el método de los momentos para θ sería $\frac{1+2+2+3+5}{5} = \frac{13}{5} = 2,6$.

Ejemplo 8.23: Sea X_1, \dots, X_n muestra aleatoria simple de $X \sim B(\theta, 1)$ con densidad

$$f(x, \theta) = \begin{cases} kx^{\theta-1} & \text{si } x \in (0, 1) \\ 0 & \text{fuera.} \end{cases}$$

Entonces

$$1 = \int_0^1 kx^{\theta-1}dx = \frac{k}{\theta},$$

de donde $k = \theta$. Si calculamos EX obtenemos

$$EX = \int_0^1 xf(x, \theta)dx = \int_0^1 x\theta x^{\theta-1}dx = \frac{\theta}{\theta+1}.$$

Por tanto, $\theta EX + EX = \theta$ ó equivalentemente $\theta(EX - 1) = -EX$ y entonces $\theta = \frac{EX}{1-EX}$ con lo que el estimador por el método de los momentos de θ es

$$\hat{\theta} = \frac{\bar{X}}{1-\bar{X}}.$$

Ejemplo 8.24: Sea X_1, \dots, X_n muestra aleatoria simple de $X \sim B(1, \theta)$ con densidad

$$f(x, \theta) = \begin{cases} k(1-x)^{\theta-1} & \text{si } x \in (0, 1) \\ 0 & \text{fuera.} \end{cases}$$

Si se calcula como antes obtenemos que $k = \theta$ y entonces

$$\begin{aligned} EX &= \int_0^1 xf(x, \theta)dx = \int_0^1 \theta x(1-x)^{\theta-1}dx \stackrel{x=1-t}{=} \theta \int_1^0 (1-t)t^{\theta-1}(-dt) \\ &= \theta \int_0^1 (1-t)t^{\theta-1}(dt) = \theta \left(\frac{1}{\theta} - \frac{1}{\theta+1} \right) = \frac{1}{\theta+1}. \end{aligned}$$

Así que, $(\theta + 1)EX = 1 \Rightarrow \theta = \frac{1-EX}{EX}$ y el estimador por el método de los momentos será

$$\hat{\theta} = \frac{1-\bar{X}}{\bar{X}}.$$

Bibliografía

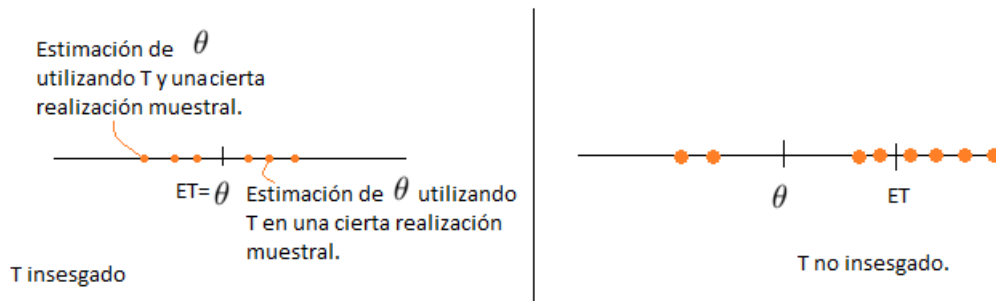
1. E.L. Lehmann (1983). *Theory of Point Estimation*. John Wiley and Sons.
2. F.M. Dekking, C. Kraaikam, H.P. Lopuhaä and L.E. Meester (2005). *A Modern Introduction to Probability and Statistics. Understanding Why and How*. Springer.
3. M.H. DeGroot (1988). *Probability and Statistics*. Wilmington, Delaware, E.U.A. Addison-Wesley Iberoamericana, S.A.
4. V.K. Rohatgi and A.K. Md. Ehsanes Saleh (2001). *An Introduction to Probability and Statistics*. John Wiley and Sons.

Tema 9

Estimación puntual II. Métodos para evaluar estimadores: error cuadrático medio, suficiencia e insesgadez.

9.1 Estimadores insesgados o centrados

Definición 9.1: Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de una población X que se distribuye según $f(x, \theta)$, θ desconocido. Diremos que el estadístico $T = T(X_1, X_2, \dots, X_n)$ es insesgado para estimar θ si $ET = \theta$.



Ejemplo 9.1: Sabemos que si X_1, X_2, \dots, X_n es una muestra aleatoria simple de X tal que $EX = \theta$, entonces $E\bar{X} = \theta$.

En particular, si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $B(1, p)$, $E\bar{X} = EX = p$, entonces \bar{X} es insesgado para estimar p .

Ejemplo 9.2: Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $Bin(m, p)$, $E\bar{X} = EX = mp$, entonces \bar{X} no es insesgado para estimar p . Pero $E(\frac{\bar{X}}{m}) = p$, y $\frac{\bar{X}}{m}$ es insesgado para estimar p .

Ejemplo 9.3: Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $P(\lambda)$, $E\bar{X} = EX = \lambda$, entonces \bar{X} es insesgado para estimar λ .

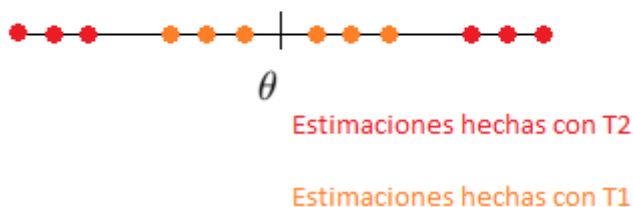
Ejemplo 9.4: Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $N(\mu, \sigma)$, $E\bar{X} = EX = \mu$, entonces \bar{X} es insesgado para estimar μ . $E(S^2) = \frac{n-1}{n}\sigma^2$, entonces S^2 no es insesgado para estimar σ^2 . Pero $E(\hat{S}^2) = \sigma^2$, y \hat{S}^2 es un estimador insesgado para σ . Sin embargo $E\hat{S} \neq \sigma$ y $ES \neq \sigma$ así que ni \hat{S} ni S son insesgados para estimar σ . Esto es así porque $E(\sqrt{T}) \neq \sqrt{ET}$.

Ejemplo 9.5: Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $U(0, \theta)$, $E\bar{X} = EX = \frac{\theta}{2}$, entonces \bar{X} no es insesgado para estimar θ . Pero $E(2\bar{X}) = \theta$, y $2\bar{X}$ es un estimador insesgado para θ .

Ejemplo 9.6: Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $\exp(\lambda)$, $E\bar{X} = EX = p$ y \bar{X} es insesgado para estimar p .

9.1.1 Eficiencia

Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de una población X que se distribuye según $f(x, \theta)$, θ desconocido. Si T_1, T_2 son dos estimadores insesgados de θ diremos que T_1 es más eficiente que T_2 si $VarT_1 < VarT_2$.



En general, dados dos estimadores insesgados de θ será preferible aquel que sea más eficiente ya que las estimaciones que proporciona están en promedio más cerca de θ .

Si dada una población y un parámetro desconocido θ , entre todos los estimadores insesgados de θ , existiera uno con menor varianza, ese sería el óptimo bajo los criterios de insesgadez y eficiencia. Bajo ciertas condiciones de regularidad es posible determinar la mínima varianza posible para los estimadores insesgados de un cierto parámetro. Tal mínima varianza se conoce como la cota Cramér-Rao. Si en la clase de los estimadores insesgados existiera uno cuya varianza coincidiera con tal cota, automáticamente tendríamos de un estimador insesgado de mínima varianza.

9.1.2 Cota de Cramér-Rao

Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de una población X que se distribuye según $f(x, \theta)$, θ desconocido. Supongamos que el soporte de X , $S = \{x | f(x, \theta) > 0\}$ no depende de θ y sea T un estimador insesgado para θ . Entonces:

$$VarT \geq \frac{1}{nE\left[\frac{\partial}{\partial\theta} \ln f(X, \theta)\right]^2}.$$

A $\frac{1}{nE\left[\frac{\partial}{\partial\theta} \ln f(X, \theta)\right]^2}$, cuando existe, se le conoce como la cota de Cramér-Rao.

Ejemplo 9.7: Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $N(\mu, \sigma)$, $S = \{x | f(x, \theta) > 0\} = \mathbb{R}$ y no depende ni de μ ni de σ .

Ejemplo 9.8: Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $P(\lambda)$, $S = \{x | f(x, \theta) > 0\} = \mathbb{N} \cup \{0\}$ y no depende de λ .

Ejemplo 9.9: Si X_1, X_2, \dots, X_n es una muestra aleatoria simple de $U(0, \theta)$, $S = \{x | f(x, \theta) > 0\} = (0, \theta)$ que depende de θ .

Por tanto, en los dos primeros casos existe la cota de Cramér-Rao, pero no en el último.

Ejemplo 9.10: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de $B(1, p)$, p desconocido. $S = \{x | f(x, \theta) > 0\} = \{0, 1\}$, que no depende de p , luego existe la cota de Cramér-Rao.

$$\begin{aligned}
 f(x, p) &= p^x(1-p)^{1-x}, \text{ si } x = 0, 1 \\
 \ln f(x, p) &= x \ln p + (1-x) \ln(1-p) \\
 \frac{\partial}{\partial p} \ln f(x, p) &= \frac{x}{p} + \frac{1-x}{1-p}(-1) = \frac{x-p}{p(1-p)} \\
 E\left[\frac{\partial}{\partial \theta} \ln f(X, \theta)\right]^2 &= E\left[\frac{X-p}{p(1-p)}\right]^2 = \frac{E(X-p)^2}{p^2(1-p)^2} = \frac{\text{Var}X}{p^2(1-p)^2} = \frac{1}{p(1-p)} \\
 C.C.R &= \frac{1}{n \frac{1}{p(1-p)}} = \frac{p(1-p)}{n} = \frac{\text{Var}X}{n} = \text{Var}\bar{X}
 \end{aligned}$$

Entonces \bar{X} es insesgado para p pues $E\bar{X} = EX = p$ y $\text{Var}\bar{X} = C.C.R$. Luego \bar{X} es estimador insesgado de mínima varianza para estimar p .

Ejemplo 9.11: Sea X_1, \dots, X_n una muestra aleatoria simple de una población X . Se pide calcular la Cota de Cramér-Rao y el Estimador Insesgado de Mínima Varianza para estimar el parámetro desconocido en los siguientes casos:

- a) $X \sim \text{Bin}(m, p)$, m conocido y p desconocido.
- b) $X \sim P(\lambda)$, λ desconocido.
- c) $X \sim N(\mu, \sigma_0)$, σ_0 conocido y μ desconocido.

a) $X \sim \text{Bin}(m, p)$.

Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de $\text{Bin}(m, p)$, m conocido y p desconocido. $S = \{x | f(x, p) > 0\} = \{0, 1, \dots, m\}$ que no depende de p , luego existe la cota de Cramér-Rao.

$$\begin{aligned}
 f(x, p) &= \binom{m}{x} p^x(1-p)^{m-x}, \text{ si } x = 0, 1, \dots, m \\
 \ln f(x, p) &= \ln \binom{m}{x} + x \ln p + (m-x) \ln(1-p) \\
 \frac{\partial}{\partial p} \ln f(x, p) &= \frac{x}{p} + \frac{1-x}{m-p}(-1) = \frac{x-mp}{p(m-p)}
 \end{aligned}$$

$$E\left[\frac{\partial}{\partial p} \ln f(X, p)\right]^2 = E\left[\frac{X - mp}{p(1-p)}\right]^2 = \frac{E(X - mp)^2}{p^2(1-p)^2} = \frac{\text{Var}X}{p^2(1-p)^2} = \frac{m}{p(1-p)}$$

$$C.C.R = \frac{1}{n \frac{m}{p(1-p)}} = \frac{p(1-p)}{nm}$$

Entonces $E\bar{X} = EX = mp$ y no es insesgado para p . Pero $E(\frac{\bar{X}}{m}) = \frac{1}{m}E\bar{X} = p$ y $\text{Var}(\frac{\bar{X}}{m}) = \frac{1}{m^2} \frac{\text{Var}X}{n} = \frac{p(1-p)}{nm} = C.C.R$. Luego $\frac{\bar{X}}{m}$ es el estimador insesgado de mínima varianza para estimar p .

b) $X \rightsquigarrow P(\lambda)$.

Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de $P(\lambda)$, λ desconocido. $S = \{x | f(x, \lambda) > 0\} = \mathbb{N} \cup \{0\}$ que no depende de λ , luego existe la cota de Cramér-Rao.

$$f(x, \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}, \text{ si } x \in \mathbb{N} \cup \{0\}$$

$$\ln f(x, \lambda) = x \ln \lambda - \ln(x!) - \lambda$$

$$\frac{\partial}{\partial \lambda} \ln f(x, \lambda) = \frac{x}{\lambda} - 1 = \frac{x - \lambda}{\lambda}$$

$$E\left[\frac{\partial}{\partial \lambda} \ln f(X, \lambda)\right]^2 = E\left[\frac{X - \lambda}{\lambda}\right]^2 = \frac{1}{\lambda^2} E(X - \lambda)^2 = \frac{1}{\lambda^2} \text{Var}(X) = \frac{1}{\lambda}$$

$$C.C.R = \frac{1}{n \frac{1}{\lambda}} = \frac{\lambda}{n} = \frac{\text{Var}(X)}{n} = \text{Var}(\bar{X})$$

Por tanto, \bar{X} es tal que $E\bar{X} = EX = \lambda$ y es insesgado para λ y $\text{Var}(\bar{X}) = C.C.R$. Luego \bar{X} es el estimador insesgado de mínima varianza.

c) $X \rightsquigarrow N(\mu, \sigma_0)$, μ desconocido y σ_0 conocido.

Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de $N(\mu, \sigma_0)$, μ desconocido. $S = \{x | f(x, \mu) > 0\} = \mathbb{R}$ que no depende de μ , luego existe la cota de Cramér-Rao.

$$f(x, \mu) = \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma_0^2}}, \text{ si } x \in \mathbb{R}$$

$$\ln f(x, \mu) = -\ln \sigma_0 - \ln(\sqrt{2\pi}) - \frac{(x-\mu)^2}{2\sigma_0^2}$$

$$\frac{\partial}{\partial \mu} \ln f(x, \mu) = \frac{-1}{2\sigma_0^2} 2(x-\mu)(-1) = \frac{x-\mu}{\sigma_0^2}$$

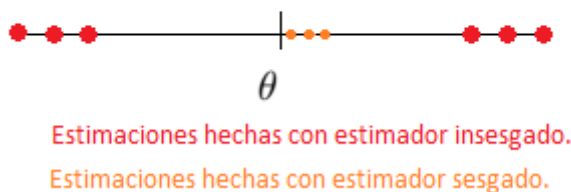
$$E\left[\frac{\partial}{\partial \mu} \ln f(X, \mu)\right]^2 = E\left[\frac{X-\mu}{\sigma_0^2}\right]^2 = \frac{1}{\sigma_0^4} E(X-\mu)^2 = \frac{1}{\sigma_0^4} \text{Var}(X) = \frac{1}{\sigma_0^2}$$

$$C.C.R = \frac{1}{n \frac{1}{\sigma_0^2}} = \frac{\sigma_0^2}{n} = \frac{Var(X)}{n} = Var(\bar{X})$$

Por tanto, \bar{X} es tal que $E\bar{X} = EX = \mu$ y es insesgado para μ y $Var(\bar{X}) = C.C.R$. Luego \bar{X} es el estimador insesgado de mínima varianza.

¿Es siempre preferible un estimador insesgado a uno sesgado?

Consideremos el siguiente ejemplo:



En él se pone de manifiesto que, en ocasiones, un estimador sesgado puede ser preferible a uno insesgado. ¿Cómo saber cuándo se tiene una situación como la anterior? Vamos a definir un criterio de comparación de estimadores que sirva para comparar estimadores cualesquiera y no solo estimadores insesgados como hace el criterio de eficiencia.

Definición 9.2: Dada una muestra aleatoria simple, X_1, \dots, X_n de $X \sim f(x, \theta)$ y $T = T(X_1, \dots, X_n)$ un estimador, definimos el error cuadrático medio de T para estimar θ como $ECM(T) = E(T - \theta)^2$.

Es claro que si T es un estimador insesgado de θ , entonces $ECM(T) = E(T - \theta)^2 = E(T - ET)^2 = Var(T)$.

Al comparar dos estimadores cualesquiera de un parámetro θ será preferible el que tenga menor error cuadrático medio, dado que en promedio, las estimaciones hechas con él estarán más cerca del verdadero valor de θ .

Por otro lado, al comparar dos estimadores insesgados utilizando el criterio del error cuadrático medio, en la práctica estaremos utilizando el criterio de eficiencia porque el ECM , en ese caso, coincide con la varianza.

Ejemplo 9.12: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de una población $X \sim U(0, \theta)$ y consideremos los siguientes estimadores de θ :

$$T = \max\{X_1, X_2, \dots, X_n\}, U = 2\bar{X} \text{ y } V = \frac{n+1}{n} \max\{X_1, X_2, \dots, X_n\}.$$

Se pide:

a) Demostrar que T no es insesgado para estimar θ pero que U y V sí que lo son.

b) Calcular el ECM de los tres estimadores.

a) Para calcular ET , necesitamos la distribución de T . Dado que T es una transformación no biyectiva de la muestra, no podemos utilizar el teorema jacobiano para calcular su distribución. Entonces calcularemos $F_T(t)$.

Si $t \in (0, \theta)$,

$$\begin{aligned} F_T(t) &= P(T \leq t) = P(\max\{X_1, X_2, \dots, X_n\} \leq t) = P(X_1 \leq t, X_2 \leq t, \dots, X_n \leq t) \\ &= \prod_{i=1}^n P(X_i \leq t) = \prod_{i=1}^n F_{U(0, \theta)} = \prod_{i=1}^n \left(\frac{t}{\theta}\right). \end{aligned}$$

Por tanto,

$$F_T(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ \frac{t^n}{\theta^n} & \text{si } t \in (0, \theta) \\ 1 & \text{si } t \geq \theta \end{cases} \Rightarrow f_T(t) = \begin{cases} \frac{nt^{n-1}}{\theta^n} & \text{si } t \in (0, \theta) \\ 0 & \text{fuera.} \end{cases}$$

Entonces

$$ET = \int t f_T(t) dt = \int_0^\theta \frac{nt^n}{\theta^n} dt = \frac{n}{\theta^n} \left[\frac{t^{n+1}}{n+1} \right]_0^\theta = \frac{n\theta}{n+1} \neq \theta,$$

por lo que T no es insesgado para estimar θ .

$$EU = E(2\bar{X}) = 2E\bar{X} = 2EX = 2E(U(0, \theta)) = 2\frac{\theta}{2} = \theta,$$

luego U es insesgado para estimar θ .

$$EV = E\left(\frac{n+1}{n}T\right) = \frac{n+1}{n}ET = \frac{n+1}{n} \frac{n\theta}{n+1} = \theta,$$

y U es insesgado para estimar θ .

b)

$$\begin{aligned} ECM(T) &= E(T - \theta)^2 = \int (t - \theta)^2 f_T(t) dt = \int_0^\theta (t - \theta)^2 \frac{nt^{n-1}}{\theta^n} dt \\ &= \frac{n}{\theta^n} \int_0^\theta (t^2 + \theta^2 - 2t\theta) t^{n-1} dt = \frac{2\theta^2}{(n+1)(n+2)}. \end{aligned}$$

$$ECM(U) = E(U - \theta)^2 = Var(U) = Var(2\bar{X}) = 4Var(\bar{X}) = 4 \frac{Var(X)}{n}$$

$$= \frac{4}{n} \text{Var}[U(0, \theta)] = \frac{4}{n} \frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

$$\begin{aligned} ECM(V) &= ECM\left(\frac{n+1}{n}T\right) = E\left(\frac{n+1}{n}T - \theta\right)^2 \\ &= \text{Var}\left(\frac{n+1}{n}T\right) = \frac{(n+1)^2}{n^2} \text{Var}(T) = \frac{(n+1)^2}{n^2} (ET^2 - E^2T), \end{aligned}$$

donde,

$$ET^2 = \int t^2 \frac{nt^{n-1}}{\theta^n} dt = \frac{n}{\theta^n} \left[\frac{t^{n+2}}{n+2} \right]_0^\theta = \frac{n\theta^2}{n+2}.$$

Entonces,

$$\frac{(n+1)^2}{n^2} (ET^2 - E^2T) = \frac{(n+1)^2}{n^2} \left(\frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} \right) = \frac{\theta^2}{n(n+2)}.$$

Ordenamos los ECM de los diferentes estadísticos y obtenemos que $\frac{\theta^2}{3n} > \frac{2\theta^2}{(n+1)(n+2)} > \frac{\theta^2}{n(n+1)}$ si $n \geq 3$. Entonces V , que es insesgado, tiene menor ECM que T que es sesgado pero T tiene menor ECM que U que es insesgado.

El procedimiento para determinar el estimador insesgado de mínima varianza a partir de la cota de Cramér-Rao es un procedimiento que no es útil en todas las ocasiones. Esto es así por dos razones:

i) Existen modelos paramétricos en los que no hay cota de Cramér-Rao. Esto sucede, por ejemplo, en el modelo $U(0, \theta)$, donde el soporte $S = \{x | f(x, \theta) > 0\} = (0, \theta)$ depende de θ y por tanto no existe la CCR para estimar θ .

ii) Existiendo la cota, el estimador insesgado de mínima varianza puede tener una varianza estrictamente superior a la cota. En tal caso no es posible determinar un estadístico con varianza igual a la cota. Esto sucede, por ejemplo, en el modelo $\exp(\lambda)$, como se verá más adelante.

Para soslayar estas excepciones a la hora de determinar el estimador insesgado de mínima varianza, se refinó el procedimiento creando uno nuevo a partir del concepto de estadístico suficiente.

9.1.3 Estadístico Suficiente

Definición 9.3: Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de una población X que se distribuye según $f(x, \theta)$, θ desconocido. Diremos que T es un estadístico suficiente para estimar θ si contiene toda la información de la muestra. Es decir, a efectos prácticos para estimar θ es igualmente útil conocer $T = T(X_1, X_2, \dots, X_n)$ que conocer X_1, X_2, \dots, X_n .

Dado que la definición anterior del estadístico suficiente es poco útil para determinar tal estadístico, Neyman y Fisher idearon su famoso criterio de factorización para establecer en cada modelo cual sería el estadístico suficiente.

Criterio de factorización de Neyman y Fisher

Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de una población X que se distribuye según $f(x, \theta)$, θ desconocido. Entonces $T = T(X_1, X_2, \dots, X_n)$ es un estadístico suficiente para estimar θ si y solo si la función de verosimilitud muestral $f(x_1, x_2, \dots, x_n, \theta)$ factoriza de la siguiente manera:

$$f(x_1, x_2, \dots, x_n, \theta) = g(t, \theta) \cdot h(x_1, x_2, \dots, x_n).$$

Siendo $g(t, \theta)$ una función de la muestra y del parámetro desconocido que depende de la muestra sólo a través de $T = t$ mientras que $h(x_1, x_2, \dots, x_n)$ es una función solo de la muestra y que, por tanto, no depende de θ .

Ejemplo 9.13: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de una población $X \sim B(1, p)$. Entonces,

$$\begin{aligned} f(x_1, x_2, \dots, x_n, p) &= \prod_{i=1}^n f(x_i, p) = \prod_{i=1}^n [p^{x_i} (1-p)^{1-x_i}] \\ &= \begin{cases} \underbrace{p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}}_{g(t, p)} \cdot h(x_1, x_2, \dots, x_n) & \text{si } x_1, x_2, \dots, x_n \in \{0, 1\} \\ 0 & \text{fuera.} \end{cases} \end{aligned}$$

Siendo

$$h(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, x_2, \dots, x_n \in \{0, 1\} \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \sum_{i=1}^n X_i$ es el estadístico suficiente. Es decir, no es necesario conocer el resultado de cada ensayo, sino sólo el número total de éxitos.

En el ejemplo se aprecia que el estadístico suficiente no es único. Cualquier transformación lineal de un estadístico suficiente proporciona otro estadístico suficiente. Trabajaremos con el más sencillo.

Ejemplo 9.14: Sea X_1, \dots, X_n una muestra aleatoria simple de una población X . Determinar un estadístico suficiente para estimar el parámetro desconocido en cada uno de los siguientes casos:

- a) p si $X \sim \text{Bin}(m, p)$, p desconocido.
- b) θ si $X \sim U(0, \theta)$, θ desconocido.
- c) λ si $X \sim \exp(\lambda)$, λ desconocido.
- d) a si $X \sim \Gamma(a, p_0)$, a desconocido.
- e) p si $X \sim \Gamma(a_0, p)$, p desconocido.
- f) μ si $X \sim N(\mu, \sigma_0)$, μ desconocido.
- g) σ si $X \sim N(\mu_0, \sigma)$, σ desconocido.
- h) p si $X \sim P(\lambda)$, λ desconocido.

- a) p si $X \sim \text{Bin}(m, p)$.

$$\begin{aligned}
f(x_1, x_2, \dots, x_n, p) &= \prod_{i=1}^n f(x_i, p) = \prod_{i=1}^n \left[\binom{m}{x_i} p^{x_i} (1-p)^{m-x_i} \right] \\
&= \begin{cases} \left[\prod_{i=1}^n \binom{m}{x_i} \right] p^{\sum_{i=1}^n x_i} (1-p)^{nm - \sum_{i=1}^n x_i} & \text{si } x_1, x_2, \dots, x_n \in \{0, 1, 2, \dots, m\} \\ 0 & \text{fuera.} \end{cases} \\
&= \underbrace{\left[\prod_{i=1}^n \binom{m}{x_i} \right] \cdot h_1(x_1, x_2, \dots, x_n)}_{h(x_1, x_2, \dots, x_n)} \cdot \underbrace{p^{\sum_{i=1}^n x_i} (1-p)^{nm - \sum_{i=1}^n x_i}}_{g(t, p)},
\end{aligned}$$

siendo

$$h_1(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, x_2, \dots, x_n \in \{0, 1, 2, \dots, m\} \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \sum_{i=1}^n X_i$ es un estadístico suficiente para estimar p .

b) θ si $X \rightsquigarrow U(0, \theta)$.

$$\begin{aligned}
f(x_1, x_2, \dots, x_n, \theta) &= \prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n \frac{1}{\theta} = \begin{cases} \frac{1}{\theta^n} & \text{si } x_1, x_2, \dots, x_n \in (0, \theta) \\ 0 & \text{fuera.} \end{cases} \\
&= \begin{cases} \frac{1}{\theta^n} & \text{si } \min\{x_1, x_2, \dots, x_n\} > 0 \text{ y } \max\{x_1, x_2, \dots, x_n\} < \theta \\ 0 & \text{fuera.} \end{cases} \\
&= \frac{1}{\theta^n} \cdot h_1(x_1, x_2, \dots, x_n) g_1(x_1, x_2, \dots, x_n, \theta),
\end{aligned}$$

siendo

$$h_1(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } \min\{x_1, x_2, \dots, x_n\} > 0 \\ 0 & \text{fuera.} \end{cases}$$

y

$$g_1(x_1, x_2, \dots, x_n, \theta) = \begin{cases} 1 & \text{si } \max\{x_1, x_2, \dots, x_n\} < \theta \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \max\{X_1, X_2, \dots, X_n\}$ es un estadístico suficiente para estimar θ .

c) λ si $X \rightsquigarrow \exp(\lambda)$.

$$\begin{aligned}
f(x_1, x_2, \dots, x_n, \lambda) &= \prod_{i=1}^n f(x_i, \lambda) = \prod_{i=1}^n (\lambda e^{-\lambda x_i}) \\
&= \begin{cases} \lambda^n e^{-\lambda \sum_{i=1}^n x_i} & \text{si } x_1, x_2, \dots, x_n \in \mathbb{R}^+ \\ 0 & \text{fuera.} \end{cases}
\end{aligned}$$

$$= \underbrace{\lambda^n e^{-\lambda \sum_{i=1}^n x_i}}_{g(t=\sum_{i=1}^n x_i, \lambda)} \cdot h_1(x_1, x_2, \dots, x_n),$$

siendo

$$h_1(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, x_2, \dots, x_n \in \mathbb{R}^+ \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \sum_{i=1}^n X_i$ es un estadístico suficiente para estimar λ .

d) a si $X \sim \Gamma(a, p_0)$, p_0 conocido.

$$\begin{aligned} f(x_1, x_2, \dots, x_n, a) &= \prod_{i=1}^n f(x_i, a) = \prod_{i=1}^n \left[\frac{p_0^a}{\Gamma(a)} e^{-p_0 x_i} x_i^{a-1} \right] \\ &= \begin{cases} \frac{p_0^{na}}{[\Gamma(a)]^n} e^{-p_0 \sum_{i=1}^n x_i} (\prod_{i=1}^n x_i)^{a-1} & \text{si } x_1, x_2, \dots, x_n \in \mathbb{R}^+ \\ 0 & \text{fuera.} \end{cases} \\ &= \underbrace{\frac{p_0^{na}}{[\Gamma(a)]^n} (\prod_{i=1}^n x_i)^{a-1}}_{g(t=\prod_{i=1}^n x_i, a)} \underbrace{e^{-p_0 \sum_{i=1}^n x_i} \cdot h_1(x_1, x_2, \dots, x_n)}_{h(x_1, x_2, \dots, x_n)}, \end{aligned}$$

siendo

$$h_1(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, x_2, \dots, x_n \in \mathbb{R}^+ \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \prod_{i=1}^n X_i$ es un estadístico suficiente para estimar a .

e) p si $X \sim \Gamma(a_0, p)$, a_0 conocido.

$$\begin{aligned} f(x_1, x_2, \dots, x_n, p) &= \prod_{i=1}^n f(x_i, p) = \prod_{i=1}^n \left[\frac{p^{a_0}}{\Gamma(a_0)} e^{-p x_i} x_i^{a_0-1} \right] \\ &= \begin{cases} \frac{p^{na_0}}{[\Gamma(a_0)]^n} e^{-p \sum_{i=1}^n x_i} (\prod_{i=1}^n x_i)^{a_0-1} & \text{si } x_1, x_2, \dots, x_n \in \mathbb{R}^+ \\ 0 & \text{fuera.} \end{cases} \\ &= \underbrace{p^{na_0} e^{-p \sum_{i=1}^n x_i}}_{g(t=\sum_{i=1}^n x_i, p)} \underbrace{\frac{(\prod_{i=1}^n x_i)^{a_0-1}}{[\Gamma(a_0)]^n} \cdot h_1(x_1, x_2, \dots, x_n)}_{h(x_1, x_2, \dots, x_n)}, \end{aligned}$$

siendo

$$h_1(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, x_2, \dots, x_n \in \mathbb{R}^+ \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \sum_{i=1}^n X_i$ es un estadístico suficiente para estimar p .

f) μ si $X \rightsquigarrow N(\mu, \sigma_0)$, σ_0 conocido.

$$\begin{aligned} f(x_1, x_2, \dots, x_n, \mu) &= \prod_{i=1}^n f(x_i, \mu) = \prod_{i=1}^n \left[\frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma_0^2}} \right] \\ &= \frac{1}{(\sigma_0 \sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma_0^2}} \\ &= \underbrace{\frac{1}{(\sigma_0 \sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i)^2}{2\sigma_0^2}}}_{h(x_1, x_2, \dots, x_n)} \cdot \underbrace{e^{-\frac{\mu^2 n}{2\sigma_0^2}} e^{\frac{\sum_{i=1}^n x_i \mu}{\sigma_0^2}}}_{g(t = \sum_{i=1}^n x_i, \mu)} \end{aligned}$$

Por el criterio de factorización $T = \sum_{i=1}^n X_i$ es el estadístico suficiente para estimar μ .

g) σ si $X \rightsquigarrow N(\mu_0, \sigma)$, μ_0 conocido.

$$\begin{aligned} f(x_1, x_2, \dots, x_n, \sigma) &= \prod_{i=1}^n f(x_i, \sigma) = \prod_{i=1}^n \left[\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_i - \mu_0)^2}{2\sigma^2}} \right] \\ &= \frac{1}{(\sigma \sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i - \mu_0)^2}{2\sigma^2}} \\ &= \underbrace{\frac{1}{(\sqrt{2\pi})^n}}_{h(x_1, x_2, \dots, x_n)} \cdot \underbrace{\frac{1}{\sigma^n} e^{-\frac{\sum_{i=1}^n (x_i - \mu_0)^2}{2\sigma^2}}}_{g(t = \sum_{i=1}^n (x_i - \mu_0)^2)} \end{aligned}$$

Por el criterio de factorización $T = \sum_{i=1}^n (X_i - \mu_0)^2$ es el estadístico suficiente para estimar σ .

h) λ si $X \rightsquigarrow P(\lambda)$.

$$\begin{aligned} f(x_1, x_2, \dots, x_n, \lambda) &= \prod_{i=1}^n f(x_i, \lambda) = \prod_{i=1}^n \left[\frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right] \\ &= \begin{cases} e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} & \text{si } x_1, x_2, \dots, x_n \in \mathbb{N} \cup \{0\} \\ 0 & \text{fuera.} \end{cases} \\ &= \underbrace{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}_{g(t = \sum_{i=1}^n x_i, \lambda)} \underbrace{\frac{1}{\prod_{i=1}^n x_i!} \cdot h_1(x_1, x_2, \dots, x_n)}_{h(x_1, x_2, \dots, x_n)} \end{aligned}$$

siendo

$$h_1(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, x_2, \dots, x_n \in \mathbb{N} \cup \{0\} \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \sum_{i=1}^n X_i$ es el estadístico suficiente para estimar λ .

Teorema 9.1: (*Rao-Blackwell*) Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de una población $X \sim f(x, \theta)$, θ parámetro desconocido. Sea U un estimador insesgado cualquiera de θ y T un estadístico suficiente para estimar θ . Definimos un nuevo estimador de θ , $V = E(U/T)$, es decir, la esperanza de la distribución condicional de U a T .

a) V es también estimador insesgado de θ .

b) $\text{Var}(V) \leq \text{Var}(U)$

a) $\text{Var}(V) = \text{Var}(U)$ si y solo si U es función de T .

Corolario: : Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de una población $X \sim f(x, \theta)$. Entonces el estimador insesgado de mínima varianza para θ es necesariamente función del estadístico suficiente.

Demostración: Supongamos que el estimador insesgado de mínima varianza es V y no es función de T , un estimador suficiente. Por reducción al absurdo, tenemos por el teorema anterior que el estimador $V = E(U/T)$ verifica:

a) Es insesgado para θ

b) $\text{Var}(V) < \text{Var}(U)$ porque por el apartado c) del teorema anterior, la igualdad no puede darse al no ser V función de T .

Entonces V no puede ser el estimador insesgado de mínima varianza porque hemos encontrado otro insesgado y con menor varianza.

A priori puede haber muchas funciones insesgadas de estadísticos suficientes.

Teorema 9.2:(*Lheman-Scheffe*) Bajo la condición de completitud (que se cumple en todos los modelos usuales) existe una única función insesgada del estadístico minimal suficiente. Dicha función es el estimador insesgado de mínima varianza.

Por tanto, para calcular el estimador insesgado de mínima varianza en un modelo concreto basta con corregir el sesgo del estadístico suficiente.

Ejemplo 9.15: Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de una población $X \sim B(1, p)$.

Entonces $T = \sum_{i=1}^n X_i$ es suficiente para estimar p .

$$ET = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n EX_i = \sum_{i=1}^n p = np \Rightarrow T \text{ no es insesgado}$$

Corregimos el sesgo: $E\left(\frac{T}{n}\right) = \frac{1}{n}ET = \frac{1}{n}np = p$, entonces $\frac{T}{n} = \bar{X}$ es función de T e insesgado para p , por lo que es el estimador insesgado de mínima varianza para p .

Ejemplo 9.16: Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de una población $X \sim \text{Bin}(m, p)$.

Entonces $T = \sum_{i=1}^n X_i$ es suficiente para estimar p .

$$ET = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n EX_i = \sum_{i=1}^n mp = nmp \Rightarrow T \text{ no es insesgado.}$$

Corregimos el sesgo $E\left(\frac{T}{nm}\right) = \frac{1}{nm}ET = \frac{1}{nm}nmp = p$, entonces $\frac{T}{nm} = \frac{\bar{X}}{m}$ es función de T e insesgado para p , por lo que es el estimador insesgado de mínima varianza para p .

Ejemplo 9.17: Sea X_1, \dots, X_n es una muestra aleatoria simple de una población X . Determinar el estimador insesgado de mínima varianza corrigiendo el sesgo del estadístico suficiente para:

- a) p si $X \sim P(\lambda)$.
- b) p si $X \sim \Gamma(a_0, p)$, a_0 conocido.
- c) σ^2 si $X \sim N(\mu_0, \sigma)$, μ_0 conocido..
- d) μ si $X \sim N(\mu, \sigma_0)$, σ_0 conocido.
- e) θ si $X \sim U(0, \theta)$.
- f) λ si $X \sim \exp(\lambda)$.
- g) a si $X \sim \Gamma(a, p_0)$, p_0 conocido.

- a) p si $X \sim P(\lambda)$.

$T = \sum_{i=1}^n X_i$ es suficiente para estimar λ .

$$ET = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n EX_i = \sum_{i=1}^n \lambda = n\lambda \Rightarrow T \text{ no es insesgado.}$$

Corregimos el sesgo: $E\left(\frac{T}{n}\right) = \frac{1}{n}ET = \frac{1}{n}n\lambda = \lambda$, entonces $\frac{T}{n} = \bar{X}$ es función de T e insesgado para λ , por lo que es el estimador insesgado de mínima varianza para λ .

- b) p si $X \sim \Gamma(a_0, p)$, a_0 conocido.

$T = \sum_{i=1}^n X_i$ es suficiente para estimar p .

$$ET = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n EX_i = \sum_{i=1}^n \frac{a_0}{p} = n \frac{a_0}{p} \Rightarrow T \text{ no es insesgado.}$$

Corregimos el sesgo: $E\left(\frac{1}{T}\right) = \int_0^\infty \frac{1}{t} f_T(t) dt = \int_0^\infty \frac{1}{t} \frac{p^{a_0}}{\Gamma(a_0)} e^{-pt} X_i^{a_0-1} = \frac{p}{na_0-1}$. Entonces $\frac{na_0-1}{T}$ es función de T e insesgado para p , por lo que es el estimador insesgado de mínima varianza para p .

c) σ^2 si $X \sim N(\mu_0, \sigma)$, μ_0 conocido.

$T = \sum_{i=1}^n (X_i - \mu_0)^2$ es suficiente para estimar σ^2 .

$$\begin{aligned} ET &= E\left(\sum_{i=1}^n (X_i - \mu_0)^2\right) = \sum_{i=1}^n (EX_i^2 + \mu_0^2 - 2EX_i\mu_0) = \sum_{i=1}^n (EX_i^2 + E\mu_0^2 - E(2X_i\mu_0)) \\ &= \sum_{i=1}^n (\sigma^2 + \mu_0^2 + \mu_0^2 - 2\mu_0\mu_0) = n\sigma^2 \Rightarrow T \text{ no es insesgado.} \end{aligned}$$

Corregimos el sesgo $E\left(\frac{T}{n}\right) = \frac{1}{n}ET = \frac{1}{n}n\sigma^2 = \sigma^2$, entonces $\frac{T}{n}$ es función de T e insesgado para σ^2 , por lo que es el estimador insesgado de mínima varianza para σ^2 .

d) μ si $X \sim N(\mu, \sigma_0)$, σ_0 conocido.

$T = \sum_{i=1}^n X_i$ es suficiente para estimar μ .

$$ET = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n EX_i = \sum_{i=1}^n \mu = n\mu \Rightarrow T \text{ no es insesgado.}$$

Corregimos el sesgo: $E\left(\frac{T}{n}\right) = \frac{1}{n}ET = \frac{1}{n}n\mu = \mu$, entonces $\frac{T}{n} = \bar{X}$ es función de T e insesgado para μ , por lo que es el estimador insesgado de mínima varianza para μ .

e) θ si $X \sim U(0, \theta)$.

$T = \max\{X_1, X_2, \dots, X_n\}$ es suficiente para estimar θ .

$$ET = E(\max\{X_1, X_2, \dots, X_n\}) = \frac{n}{n+1}\theta \Rightarrow T \text{ no es insesgado.}$$

Corregimos el sesgo $E\left(\frac{n+1}{n}T\right) = \frac{n+1}{n} \frac{n}{n+1}\theta = \theta$, entonces $\frac{n+1}{n}T$ es función de T e insesgado para θ , por lo que es el estimador insesgado de mínima varianza para θ .

f) λ si $X \sim \exp(\lambda)$.

$T = \sum_{i=1}^n X_i$ es suficiente para estimar λ .

$$ET = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n EX_i = \sum_{i=1}^n \frac{1}{\lambda} = \frac{n}{\lambda} \Rightarrow T \text{ no es insesgado.}$$

Corregimos el sesgo

$$E\left(\frac{1}{T}\right) = \int_0^\infty \frac{1}{t} f_T(t) dt = \int_0^\infty \frac{1}{t} \frac{\lambda^n}{\Gamma(n)} e^{-\lambda t} t^{n-1} dt = \frac{\lambda^n}{\Gamma(n)} \int_0^\infty e^{-\lambda t} t^{n-2} dt = \frac{\lambda^n}{\Gamma(n)} \frac{\Gamma(n-1)}{\lambda^{n-1}} = \frac{\lambda}{n-1}.$$

Entonces $\frac{n-1}{T}$ es función de T e insesgado para λ , por lo que es el estimador insesgado de mínima varianza para λ .

En este caso se puede comprobar que $\frac{n-1}{T}$ no alcanza la cora de Cramér-Rao.

$$\begin{aligned} E\left(\frac{1}{T^2}\right) &= \int_0^\infty \frac{1}{t^2} f_T(t) dt = \int_0^\infty \frac{1}{t^2} \frac{\lambda^n}{\Gamma(n)} e^{-\lambda t} t^{n-1} dt = \frac{\lambda^n}{\Gamma(n)} \int_0^\infty e^{-\lambda t} t^{n-3} dt \\ &= \frac{\lambda^n}{\Gamma(n)} \frac{\Gamma(n-2)}{\lambda^{n-2}} = \frac{\lambda^2}{(n-1)(n-2)}, \end{aligned}$$

de donde

$$\begin{aligned} \text{Var}\left(\frac{n-1}{T}\right) &= (n-1)^2 \text{Var}\left(\frac{1}{T}\right) = (n-1)^2 \left[E\left(\frac{1}{T}\right)^2 - E^2\left(\frac{1}{T}\right) \right] \\ &= (n-1)^2 \left[\frac{\lambda^2}{(n-1)(n-2)} - \frac{\lambda^2}{(n-1)^2} \right] = \frac{\lambda^2}{n-2}. \end{aligned}$$

Y $\frac{\lambda^2}{n-2}$ es estrictamente mayor que $\frac{\lambda^2}{n}$, cota de Cramér-Rao en este modelo, como el lector puede comprobar.

g) a si $X \sim \Gamma(a, p_0)$, p_0 conocido.

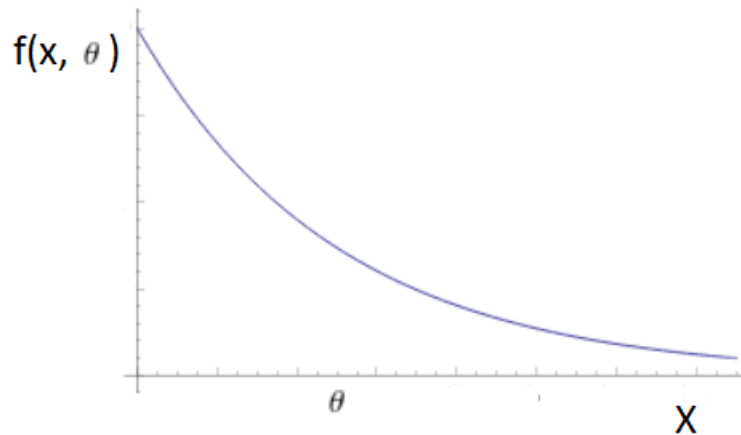
$T = \prod_{i=1}^n X_i$ es suficiente para estimar a .

$$ET = E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n EX_i = \prod_{i=1}^n \frac{a}{p_0} = \frac{a^n}{p_0^n}.$$

Aquí no sabemos como corregir el sesgo de T y no podemos obtener el estimador insesgado de mínima varianza.

Ejemplo 9.18: Sea X_1, \dots, X_n muestra aleatoria simple de $X \sim f(x, \theta)$ siendo:

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \text{si } x \geq 0 \\ 0 & \text{fuera} \end{cases}$$



Determinar:

- a) El estimador insesgado de mínima varianza.
- b) ¿Alcanza la Cota de Cramér-Rao?

- a) El estimador insesgado de mínima varianza.

$$f(x_1, \dots, x_n, \theta) = \underbrace{\frac{1}{\theta^n} e^{-\frac{\sum_{i=1}^n x_i}{\theta}}}_{g(t=\sum_{i=1}^n x_i, \lambda)} \cdot h(x_1, x_2, \dots, x_n)$$

Siendo

$$h(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } x_1, x_2, \dots, x_n \geq 0 \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \sum_{i=1}^n X_i$ es el estadístico suficiente para estimar θ .

$$ET = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = n\theta,$$

$$E\left(\frac{T}{n}\right) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{n\theta}{n} = \theta$$

Por lo que estimador insesgado de mínima varianza es \bar{X}

- b) Alcanza la Cota de Cramér-Rao?

$$C.C.R = \frac{1}{n \cdot E\left[\frac{\partial}{\partial \theta} \ln f(x, \theta)\right]^2}$$

$$f(x, \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \ln f(x, \theta) = -\ln \theta - \frac{x}{\theta},$$

$$\frac{\partial}{\partial \theta} \ln f(x, \theta) = -\frac{1}{\theta} + \frac{x}{\theta^2},$$

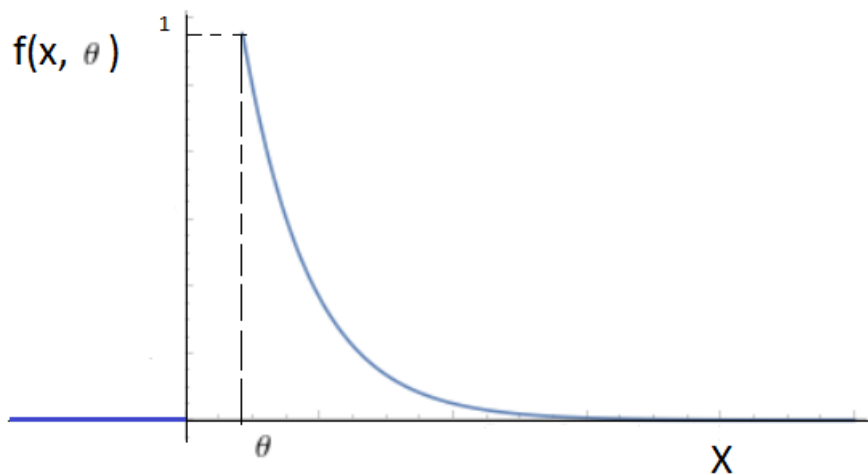
$$\begin{aligned} E\left[\frac{\partial}{\partial \theta} \ln f(X, \theta)\right]^2 &= E\left[-\frac{1}{\theta} + \frac{X}{\theta^2}\right]^2 = E\left[\frac{-\theta + X}{\theta^2}\right]^2 \\ &= \frac{E(X - \theta)^2}{\theta^4} = \frac{1}{\theta^4} E[X - \theta]^2 = \frac{1}{\theta^4} E[X - EX]^2 = \frac{1}{\theta^4} \text{Var}(X) = \frac{1}{\theta^4} \theta^2 = \frac{1}{\theta^2}, \end{aligned}$$

$$C.C.R = \frac{1}{n \cdot \frac{1}{\theta^2}} = \frac{\theta^2}{n}.$$

$$\text{Y } \text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n} = \frac{\theta^2}{n} = C.C.R$$

Ejemplo 9.19: Sea X_1, \dots, X_n muestra aleatoria simple de $X \sim f(x, \theta)$ siendo:

$$f(x, \theta) = \begin{cases} e^{-x} e^{\theta} & \text{si } x \geq \theta \\ 0 & \text{fuera} \end{cases}$$



Determinar el estimador insesgado de mínima varianza.

$$f(x_1, \dots, x_n, \theta) = \underbrace{\frac{e^{n\theta}}{g_1(x_1, x_2, \dots, x_n)}}_{g(t=\min\{x_1, x_2, \dots, x_n\}, \theta)} \underbrace{e^{-\sum_{i=1}^n x_i}}_{h(x_1, x_2, \dots, x_n)}$$

Siendo

$$g_1(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{si } \min\{x_1, x_2, \dots, x_n\} \geq \theta \\ 0 & \text{fuera.} \end{cases}$$

Por el criterio de factorización $T = \min\{X_1, X_2, \dots, X_n\}$ es el estadístico suficiente para estimar θ .

La función de distribución de T viene dada por:

$$\begin{aligned} F_T(t) &= P(T \leq t) = 1 - P(T > t) = 1 - P(\min\{X_1, X_2, \dots, X_n\} > t) \\ &= 1 - \prod_{i=1}^n P(x_i > t) = 1 - \prod_{i=1}^n \int_t^{\infty} e^{-x} e^{\theta} dx = 1 - \prod_{i=1}^n [-e^{-x}]_t^{\infty} \\ &= 1 - \prod_{i=1}^n e^{\theta} (e^{-t}) = \begin{cases} 1 - e^{n\theta} e^{-nt} & \text{si } t \geq 0 \\ 0 & \text{fuera,} \end{cases} \end{aligned}$$

de donde la densidad de T es:

$$f_T(t) = F'_T(t) = \begin{cases} ne^{n\theta} e^{-nt} & \text{si } t \geq 0 \\ 0 & \text{fuera.} \end{cases}$$

$$\begin{aligned} ET &= \int_{\theta}^{\infty} t n e^{n\theta} e^{-nt} dt = n e^{n\theta} \int_{\theta}^{\infty} t e^{-nt} dt \stackrel{u=t, du=dt}{=} \int_{dv=e^{-nt}, v=\frac{1}{n}e^{-nt}}^{\infty} t e^{-nt} dt \\ &= n e^{n\theta} \left[t \left(\frac{-1}{n} e^{-nt} \right) \right]_{\theta}^{\infty} + \int_{\theta}^{\infty} \frac{1}{n} e^{-nt} dt = \theta + \frac{1}{n} \end{aligned}$$

T no es insesgado pero $T_2 = T - \frac{1}{n}$ tiene como esperanza $ET_2 = E(T - \frac{1}{n}) = ET - E(\frac{1}{n}) = \theta + \frac{1}{n} - \frac{1}{n} = \theta$. Luego T_2 es insesgado para θ y función del estadístico suficiente, y entonces es el estimador insesgado de mínima varianza.

9.1.4 Relaciones entre estimadores.

i) El estimador máximo verosímil es siempre función del estadístico suficiente. Por tanto, cuando sea insesgado será automáticamente el estimador insesgado de mínima varianza. Este es el caso, por ejemplo, en el modelo $B(1, p)$, $Bin(m, p)$, $P(\lambda)$ y $N(\mu, \sigma_0)$, con σ_0 conocido. También se alcanza la Cota Cramér-Rao.

ii) Cuando el estimador máximo verosímil no es insesgado, corrigiendo su sesgo (ya que es función del suficiente) se obtendrá el estimador insesgado de mínima varianza. En estas situaciones no se alcanza la cota de Cramér-Rao.

Ejemplo 9.20: En el modelo $U(0, \theta)$, donde recordamos que $\hat{\theta} = \max\{X_1, \dots, X_n\}$ que no es insesgado, corrigiendo el sesgo tenemos $\frac{n+1}{n} \max\{X_1, \dots, X_n\}$ que es el estimador insesgado de mínima varianza. En este caso no se alcanza la cota, pues ésta, ni siquiera existe.

Ejemplo 9.21: En el modelo $\exp(\lambda)$ el estimador máximo verosímil es $\frac{1}{\bar{X}}$ que no es insesgado. Corrigiendo el sesgo se obtiene $\frac{n-1}{\sum_{i=1}^n x_i}$ que es el estimador insesgado de mínima varianza. Aquí aunque existe, no se alcanza la cota, como se vio.

iv) Cuando el estimador máximo verosímil se obtiene como raíz de la ecuación de verosimilitud, entonces es consistente y asintóticamente normal (CAN). Además es asintóticamente insesgado y eficiente, es decir, si $n \rightarrow \infty$, $\hat{\theta}$ sigue aproximadamente una distribución

$$N(\theta, \sqrt{\frac{1}{nE\left[\frac{\partial}{\partial\theta}\ln(f(X, \theta))^2\right]}})$$

Ejemplo 9.22: Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de $X \sim B(1, p)$. El estimador máximo verosímil \bar{X} se obtiene como raíz de la ecuación de verosimilitud, luego es CAN. Además no sólo es asintóticamente insesgado y eficiente, sino que es insesgado y alcanza la cota para cada n . Por tanto, por ser consistente, $\bar{X} \xrightarrow{P} p$, y si n es grande $\bar{X} \rightsquigarrow_{aprox} N(p, \frac{p(1-p)}{n}) \Leftrightarrow \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow_{aprox} N(0, 1)$

Ejemplo 9.23: Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de $X \sim Bin(m, p)$. El estimador máximo verosímil $\frac{\bar{X}}{m}$ se obtiene como raíz de la ecuación de verosimilitud, luego es CAN. Además no sólo es asintóticamente insesgado y eficiente, sino que es insesgado y alcanza la cota para cada n . Por tanto, por ser consistente, $\frac{\bar{X}}{m} \xrightarrow{P} p$, y si n es grande $\frac{\bar{X}}{m} \rightsquigarrow_{aprox} N(p, \frac{p(1-p)}{mn})$.

Ejemplo 9.25: Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de $X \sim P(\lambda)$, el estimador máximo verosímil \bar{X} se obtiene como raíz de la ecuación de verosimilitud, luego es CAN. Además no sólo es asintóticamente insesgado y eficiente, sino que es insesgado y alcanza la cota para cada n . Por tanto, por ser consistente, $\bar{X} \xrightarrow{P} \lambda$, y si n es grande $\bar{X} \rightsquigarrow_{aprox} N(\lambda, \frac{\lambda}{n}) \Leftrightarrow \frac{\bar{X}-\lambda}{\sqrt{\frac{\lambda}{n}}} \rightsquigarrow_{aprox} N(0, 1)$.

Ejemplo 9.26: Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de $X \sim U(0, \theta)$. El estimador máximo verosímil no se obtiene como raíz de la ecuación de verosimilitud. Sin embargo, en este caso sigue siendo consistente.

$$\max\{X_1, X_2, \dots, X_n\} \xrightarrow{P} \theta.$$

Pero no es asintóticamente normal. El $\max\{X_1, X_2, \dots, X_n\}$ si n crece no tiene distribución normal, como sabemos.

Ejemplo 9.27: Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de $X \sim \exp(\lambda)$, el estimador máximo verosímil $\hat{\lambda} = \frac{1}{\bar{X}}$ se obtiene como raíz de la ecuación de verosimilitud, luego es consistente, $\frac{1}{\bar{X}} \xrightarrow{P} \lambda$. Además es asintóticamente normal, si n crece, $\frac{1}{\bar{X}} \rightsquigarrow_{aprox} N(\lambda, \sqrt{\frac{\lambda^2}{n}}) \equiv N(\lambda, \frac{\lambda}{\sqrt{n}})$.

Ejemplo 9.28: Sea X_1, X_2, \dots, X_n es una muestra aleatoria simple de $X \sim N(\mu, \sigma)$, entonces $\hat{\mu} = \bar{X}$ es el estimador máximo verosímil y se obtiene como raíz de la ecuación de verosimilitud,

luego es consistente $\bar{X} \xrightarrow{P} \mu$ y $S^2 \xrightarrow{P} \sigma^2$. Además es asintóticamente normal, es decir, si n crece $\bar{X} \rightsquigarrow_{\text{aprox}} N(\mu, \sqrt{\frac{\sigma^2}{n}}) \equiv N(\mu, \frac{\sigma}{\sqrt{n}})$.

Bibliografía

1. E.L. Lehmann (1983). *Theory of Point Estimation*. John Wiley and Sons.
2. F.M. Dekking, C. Kraaikam, H.P. Lopuhaä and L.E. Meester (2005). *A Modern Introduction to Probability and Statistics. Understanding Why and How*. Springer.
3. M.H. DeGroot (1988). *Probability and Statistics*. Wilmington, Delaware, E.U.A. Addison-Wesley Iberoamericana, S.A.
4. V.K. Rohatgi and A.K. Md. Ehsanes Saleh (2001). *An Introduction to Probability and Statistics*. John Wiley and Sons.

Tema 10

Tests de hipótesis. Introducción. Métodos para encontrar tests: tests de ratios de verosimilitud. Métodos para evaluar tests: probabilidad de error y función de potencia, tests más potentes

10.1 Tests de hipótesis

10.1.1 Introducción

Dentro de la inferencia estadística, un contraste de hipótesis (también conocido como test de hipótesis o prueba de significación) es un procedimiento para juzgar si una propiedad que se supone en una población estadística es compatible con lo observado en una muestra de dicha población. Este tipo de análisis fue iniciado por Ronald Fisher (1890-1962), y fundamentado posteriormente por Jerzy Neyman (1894-1981) y Karl Pearson (1857-1936).

Mediante esta teoría, se aborda el problema estadístico considerando una hipótesis determinada H_0 , y una hipótesis alternativa H_1 , y se trata de determinar si los datos muestran o no evidencia en contrar de H_0 .

La noción de contraste de hipótesis está fuertemente asociada al concepto estadístico de potencia y a los conceptos de errores de tipo I y II, que definen respectivamente, la posibilidad de tomar un suceso verdadero como falso, o uno falso como verdadero.

10.1.2 Motivación

Se sospecha que una moneda se ha trucado para obtener más caras que cruces cuando se lanza al aire. Para intentar demostrar esta idea intuitiva, la forma de proceder es llevar a cabo una serie de lanzamientos, por ejemplo, 30, y tomar nota de la cantidad de caras obtenidas. Si el valor es *alto*, por ejemplo, 25 o más, se considera que el resultado es poco compatible con la hipótesis de que la moneda no está trucada. En este escenario, se concluiría diciendo que las observaciones contradicen esta hipótesis.

Al aplicar ciertas técnicas probabilísticas, se puede determinar a partir de qué valor de un cierto estadístico o de una cierta probabilidad se debe rechazar la hipótesis, garantizando que la probabilidad de cometer un error es un valor conocido a priori. Las hipótesis pueden clasificarse en dos grupos, atendiendo a su naturaleza:

- Si especifican un valor concreto o un intervalo para los parámetros del modelo. Por ejemplo, la hipótesis de que la media de una variable es 10
- Si determinan el tipo de distribución de probabilidad que ha generado los datos. Por ejemplo, la distribución de probabilidad es la distribución normal.

La metodología de ambos tipos de contrastes es análoga. No obstante, es importante distinguir en qué escenario se encuadra cada problema, ya que, muchos problemas de contraste de hipótesis

respecto a un parámetro son, en realidad, problemas de estimación, que tienen una respuesta complementaria en forma de intervalo de confianza. Las hipótesis respecto a la forma de la distribución se suelen utilizar para validar un modelo estadístico para un fenómeno aleatorio que se está estudiando.

10.2 Planteamiento clásico del contraste de hipótesis

Para trabajar con contrastes de hipótesis, hay que establecer una región de rechazo. Se suele tomar el conjunto de valores que es más improbable bajo la hipótesis, o lo que es lo mismo, el conjunto de valores para el que rechazaremos la hipótesis nula. Tomando una muestra de la población en estudio, se calcula un estadístico, que es una función de la muestra. Es importante que la distribución de probabilidad de este estadístico esté relacionada con la hipótesis en estudio y sea conocida. Después se calcula la probabilidad de que un valor del estadístico entre en la región de rechazo. Así se puede establecer esta región para que la probabilidad de mantener H_0 cuando sea falsa, sea lo suficientemente pequeña.

Volviendo al ejemplo de la moneda trucada, se considera que la muestra de la población son los treinta lanzamientos realizados; el estadístico que se calcula es el número total de caras obtenidas, y la región de rechazo está definida por las muestras en las que el total de caras es igual o superior a 25. La probabilidad de cometer el error de admitir que la moneda está trucada, pese a que no lo está, es igual a la probabilidad binomial de tener 25 éxitos o más en una serie de 30 ensayos de Bernoulli con probabilidad de éxito 0,5 en cada uno, es decir, 0,0002. Esto nos dice que, aunque sea poco probable, existe la posibilidad de obtener más de 25 caras en la muestra (30 lanzamientos) pese a que la moneda no esté trucada.

Por tanto, se conoce como hipótesis nula H_0 , a la hipótesis que se desea contrastar. El nombre de “nula” significa “sin valor, efecto o consecuencia”. Por tanto, entendemos que H_0 debe identificarse con la hipótesis de no cambio (a partir de la opinión actual); no diferencia, no mejora, etc. H_0 representa la hipótesis que mantendremos a no ser que los datos indiquen su falsedad, y puede entenderse, por tanto, en el sentido de “neutra”. Es importante tener en cuenta que la hipótesis H_0 nunca se considera probada, aunque pudiera ser rechazada por los datos. Por ejemplo, la hipótesis de que dos poblaciones tienen la misma media puede ser rechazada fácilmente cuando ambas difieren mucho, analizando muestras suficientemente grandes de ambas poblaciones, pero no puede ser “demostrada” aplicando técnicas de muestreo, ya que siempre existe la posibilidad de que las medias se diferencien en una cantidad lo suficientemente pequeña para no ser detectada, aún cuando la muestra fuera muy grande.

Definición 10.1:

Sea X_1, X_2, \dots, X_n una muestra aleatoria simple (m.a.s.) de una población $X \sim f(x, \theta)$, siendo θ un parámetro desconocido. Un contraste o test de hipótesis sobre θ es una regla de decisión que permite decidir o elegir entre dos hipótesis:

- $H_0: \theta \in \Theta_0$ **Hipótesis nula.**

- $H_1: \theta \in \Theta_1$ **Hipótesis alternativa.**

siendo Θ_0 y Θ_1 conjuntos disjuntos del espacio paramétrico Θ , o lo que es lo mismo, del espacio de valores posibles para θ .

Al llevar a cabo un contraste de hipótesis, la hipótesis nula no se considera probada. Se trata de determinar si los datos presentan evidencia en contra de tal hipótesis: cuando una hipótesis nula no se rechaza, generalmente, otras hipótesis nulas similares tampoco se rechazan.

Ejemplo 10.1:

Si se lanza 100 veces una moneda y se obtienen 51 caras. En el siguiente contraste:

- $H_0: p = \frac{1}{2}$
- $H_1: p \neq \frac{1}{2}$

no será rechazada H_0 , y tampoco lo será en el contraste

- $H_0: p = \frac{51}{100}$
- $H_1: p \neq \frac{51}{100}$

Cuando los datos permiten rechazar H_0 se dice que son significativos. En cambio, si no muestran evidencia en contra de H_0 , se dice que no son significativos.

10.3 Tipos de hipótesis

Dependiendo de la forma de Θ_0 y Θ_1 , existen varios tipos de hipótesis:

- **Hipótesis simples:** son aquellas que se refieren a un único valor. Por tanto, Θ_0 y Θ_1 son ambos unipuntuales.
- **Hipótesis compuestas:** son aquellas hipótesis que no son simples. Es decir, Θ_0 y Θ_1 contienen ambos más de un valor.
- **Hipótesis unilateral:** se corresponde con todos los valores del parámetros mayores o iguales que uno dado (unilateral a la derecha), o menor o igual que uno dado (unilateral a la izquierda).
- **Hipótesis bilateral:** aquella que contiene todos los valores a ambos lados de uno dado.

Es importante tener en cuenta que las hipótesis unilaterales y bilaterales son compuestas.

Observación.

La hipótesis más relevante para el investigador es la que se coloca como hipótesis nula, tratando de determinar si los datos muestran evidencia en contra de ella.

Para obtener una regla de decisión, se debe definir lo que se conoce como región crítica “ C ” del contraste.

La región crítica de un contraste

$$\begin{cases} H_0, & \theta \in \Theta_0 \\ H_1, & \theta \in \Theta_1 \end{cases}$$

es

$$C = \{X_1, X_2, \dots, X_n \mid \text{se rechaza } H_0\}$$

es decir, el conjunto de resultados para los cuales se rechaza H_0 .

Ejemplo 10.2:

(a) Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim P(\lambda)$. Se desea contrastar

$$\begin{cases} H_0, & \lambda = 1 & \text{hipótesis simple} \\ H_1, & \lambda > 1 & \text{hipótesis compuesta} \end{cases}$$

Una posible región crítica es $C = \{X_1, X_2, \dots, X_n \mid \bar{X} > 1,7\}$, consistente en rechazar H_0 si el promedio supera el valor 1,7.

A pesar de la notación estándar que expresa la región crítica en términos de la muestra, C se debe interpretar como el conjunto de realizaciones muestrales (observaciones) cuya media supera el valor 1,7. La notación, por tanto, utilizará variables y realizaciones de dichas variables de manera indistinta.

(b) Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim N(\mu, \sigma_0)$, con media μ desconocida y desviación típica σ_0 conocida. Se desea contrastar

$$\begin{cases} H_0, & \mu = 0 & \text{hipótesis simple} \\ H_1, & \mu \neq 0 & \text{hipótesis compuesta} \end{cases}$$

Una posible región crítica en este caso es $C = \{X_1, X_2, \dots, X_{20} \mid |\bar{X}| > 0,6\}$.

(c) Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim B(1, p)$, con p desconocido. Se desea contrastar

$$\begin{cases} H_0, & p \leq \frac{1}{2} & \text{hipótesis compuesta unilateral izquierda} \\ H_1, & p > \frac{1}{2} & \text{hipótesis compuesta unilateral derecha} \end{cases}$$

Una posible región crítica en este caso es $C = \{X_1, X_2, \dots, X_{20} \mid \sum_{i=1}^{20} X_i \geq 16\}$, siendo $\sum_{i=1}^n X_i$ el número de éxitos.

Precisamente, la teoría de los contrastes de hipótesis va encaminada a determinar en cada problema la región crítica que es óptima en algún sentido.

10.4 Contrastes de hipótesis simples

Diremos que un contraste es de hipótesis simples cuando las dos hipótesis son de tipo simple.

$$\begin{cases} H_0, & \theta = \theta_0 \\ H_1, & \theta = \theta_1 \end{cases}$$

Sea una regla de decisión para este contraste, y sea C la correspondiente región crítica. Esta regla de decisión conlleva dos tipos de error:

- **Error de tipo I:** se da cuando se rechaza H_0 pese a que es cierta.
- **Error de tipo II:** se da cuando se mantiene o acepta H_0 pese a que es falsa.

Llamaremos α a la probabilidad de error de tipo I, y β a la probabilidad de error de tipo II. Entonces,

$$\alpha = p(C|H_0) := p_{H_0}(C) = p_{\theta_0}(C) \quad (10.1)$$

$$\beta = p(\bar{C}|H_1) := p_{H_1}(\bar{C}) = p_{\theta_1}(\bar{C}) \quad (10.2)$$

No es posible reducir simultáneamente ambos tipos de error.

La teoría de los contrastes de hipótesis simples fija una probabilidad máxima, en general, $\alpha = 0,05$ para el error de tipo I, y entre todas las regiones críticas que satisfacen esta condición, se elige como óptima la que minimice el valor de β .

Ejemplo 10.3:

Sea X_1, X_2, \dots, X_n una m.a.s. de $N(\mu, 2)$. Se desea contrastar las siguientes hipótesis simples:

$$H_0 : \mu = 2 \quad \text{frente a} \quad H_1 : \mu = 3$$

Si se utiliza la región crítica $C = \{X_1, X_2, \dots, X_n \mid \bar{X} > 2,87\}$, calcular la probabilidad de error tipo I y la probabilidad de error tipo II.

$$\alpha = p(\text{rechazar } H_0 \text{ siendo cierta}) = p_{H_0}(C) := p_{\mu=2}(C) = p\left(Z > \frac{2,87 - 2}{\frac{2}{\sqrt{n}}}\right) =$$

$$= p(Z > 0,435\sqrt{n}) = \begin{cases} p(Z > 1,305) = 1 - 0,91149 = 0,0851, & \text{si } n = 9 \\ p(Z > 4,35) = 0, & \text{si } n = 100 \end{cases}$$

$$\beta = p(\text{error tipo II}) = p(\text{mantener } H_0 \text{ siendo falsa}) = p_{H_1}(\bar{C}) = p_{\mu=3}(\bar{C}) = p_{\mu=3}(\bar{X} \leq 2,87) =$$

$$\begin{aligned} &=_{\bar{X} \sim N(3, \frac{2}{\sqrt{n}})} p\left(N\left(3, \frac{2}{\sqrt{n}}\right) \leq 2,87\right) = p\left(Z \leq \frac{2,87 - 3}{\frac{2}{\sqrt{n}}}\right) = p(Z \leq 0,065\sqrt{n}) \\ &= \begin{cases} p(Z \leq -0,195) = 1 - 0,57535 = 0,42465, & \text{si } n = 9 \\ p(Z \leq -0,65) = 1 - 0,74215 = 0,25785 & \text{si } n = 100 \end{cases} \end{aligned}$$

donde, para calcular las probabilidades de los distintos tipos de error, $n = 9$ y $n = 100$ son valores de tamaño muestral cualesquiera. Se podrían haber elegido otros. Lógicamente, la información, el aumento del tamaño muestral, reduce la probabilidad de error.

10.4.1 Determinación de regiones críticas óptimas en contrastes de hipótesis simples

A continuación se explica cómo minimizar una combinación lineal de las probabilidades ambos tipos de error.

Proposición 1.

Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim f(x, \theta)$, y se desea contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$, ambas hipótesis simples. Sean $f(x_1, x_2, \dots, x_n, \theta_0)$ y $f(x_1, x_2, \dots, x_n, \theta_1)$ las verosimilitudes muestrales bajo θ_0 y θ_1 , respectivamente.

Consideramos la regla de decisión δ^* con región crítica

$$C^* = \{X_1, X_2, \dots, X_n \mid af(X_1, X_2, \dots, X_n, \theta_0) < bf(X_1, X_2, \dots, X_n, \theta_1), a, b, \in \mathbb{R}\}.$$

Entonces, esta regla de decisión es tal que, para otra regla δ con región crítica C se tiene que $a\alpha(\delta^*) + b\beta(\delta^*) < a\alpha(\delta) + b\beta(\delta)$; es decir, la regla δ^* minimiza la combinación lineal de las probabilidades de error con pesos a y b .

Demostración.

Vamos a demostrarlo en el caso discreto. Para el caso absolutamente continuo, la prueba es similar.

$$\begin{aligned} a\alpha(\delta^*) + b\beta(\delta^*) &= ap_{H_0}(C^*) + bp_{H_1}(\bar{C}^*) = ap_{H_0}(C^*) + b[1 - p_{H_1}(C^*)] = \\ &a \sum_{(x_1, x_2, \dots, x_n) \in C^*} f(x_1, x_2, \dots, x_n, \theta_0) + b \left[1 - \sum_{(x_1, x_2, \dots, x_n) \in C^*} f(x_1, x_2, \dots, x_n, \theta_1) \right] = \end{aligned}$$

$$b + \sum_{(x_1, x_2, \dots, x_n) \in C^*} [af(x_1, x_2, \dots, x_n, \theta_0) - bf(x_1, x_2, \dots, x_n, \theta_1)]$$

Sea C una región crítica asociada a una regla $\delta \neq \delta^*$. Entonces, $C \neq C^*$, puesto que en C^* están todas las realizaciones (x_1, x_2, \dots, x_n) para las que $af(X_1, X_2, \dots, X_n, \theta_0) < bf(X_1, X_2, \dots, X_n, \theta_1) \iff af(X_1, X_2, \dots, X_n, \theta_0) - bf(X_1, X_2, \dots, X_n, \theta_1) < 0$.

Cualquier otra región crítica $C \neq C^*$ tendrá al menos una realización muestral (x_1, x_2, \dots, x_n) para la que $af(x_1, x_2, \dots, x_n, \theta_0) - bf(x_1, x_2, \dots, x_n, \theta_1) > 0$. Entonces, $a\alpha(\delta) + b\beta(\delta) > a\alpha(\delta^*) + b\beta(\delta^*)$.

Lema de Neyman-Pearson

Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim f(x, \theta)$, con θ desconocido. Se desea contrastar

$$H_0 : \theta = \theta_0 \quad \text{frente a} \quad H_1 : \theta = \theta_1$$

Fijado $\alpha_0 \in [0, 1]$, la regla de decisión δ^* con región crítica $C^* = \{x_1, x_2, \dots, x_n \mid f(x_1, x_2, \dots, x_n, \theta_0) < kf(x_1, x_2, \dots, x_n, \theta_1)\}$ satisface que para cualquier otra regla de decisión δ con $\alpha(\delta) \leq \alpha_0$, entonces $\beta(\delta) \geq \beta(\delta^*)$. La constante k se determina con la condición $\alpha(\delta^*) = \alpha_0$.

Como consecuencia, la regla de decisión δ^* es la que minimiza la probabilidad de error tipo II, entre todas las reglas que tienen una probabilidad de error de tipo I inferior o igual a α_0 . Usualmente, α_0 es un valor pequeño, conocido como nivel de simplificación. Por defecto, $\alpha_0 = 0,05$.

Demostración

Utilizando la proposición anterior con $a = 1$ y $b = k$, δ^* es la regla de decisión que minimiza $\alpha(\delta) + k\beta(\delta)$, entre todas las reglas de decisión δ . Sea ahora una regla δ con $\alpha(\delta) \leq \alpha_0$.

$$\left. \begin{aligned} \alpha(\delta^*) + k\beta(\delta^*) &\leq \alpha(\delta) + k\beta(\delta) \\ \alpha(\delta) &\leq \alpha_0 \\ \alpha(\delta^*) &= \alpha_0 \end{aligned} \right\}$$

Entonces, por la proposición anterior,

$$\beta(\delta) \geq \beta(\delta^*).$$

Ejemplo 10.4:

Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim N(\mu, \sigma_0)$, con media μ desconocida y desviación típica σ_0 conocida. Se desea contrastar:

$$H_0 : \mu = 0 \quad \text{frente a} \quad H_1 : \mu = 1$$

Con nivel máximo de probabilidad de error de tipo I igual a α , determinar la región crítica óptima.

Por el lema de Neyman-Pearson, y utilizando notación de realizaciones muestrales, es decir, x_1, \dots, x_n , dicha región es

$$C = \{x_1, x_2, \dots, x_n \mid f(x_1, x_2, \dots, x_n, \mu_0) < k f(x_1, x_2, \dots, x_n, \mu_1)\}$$

determinándose k para que $\alpha = p_{H_0}(C)$.

$$\begin{aligned} f(x_1, x_2, \dots, x_n, 0) < k f(x_1, x_2, \dots, x_n, 1) &\iff \frac{f(x_1, x_2, \dots, x_n, 1)}{f(x_1, x_2, \dots, x_n, 0)} > \frac{1}{k} \iff \frac{\prod_{i=1}^n f(x_i, 1)}{\prod_{i=1}^n f(x_i, 0)} > \frac{1}{k} \\ &\iff \frac{\prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}\sigma_0} e^{\frac{-1}{2\sigma_0^2}(x_i-1)^2} \right]}{\prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}\sigma_0} e^{\frac{-1}{2\sigma_0^2}x_i^2} \right]} > \frac{1}{k} \iff \frac{e^{\frac{-1}{2\sigma_0^2}\sum_{i=1}^n (x_i-1)^2}}{e^{\frac{-1}{2\sigma_0^2}\sum_{i=1}^n x_i^2}} > \frac{1}{k} \iff \frac{e^{\frac{-1}{2\sigma_0^2}\sum_{i=1}^n (x_i^2+1-2x_i)}}{e^{\frac{-1}{2\sigma_0^2}\sum_{i=1}^n x_i^2}} > \frac{1}{k} \iff \\ &\frac{e^{\frac{-1}{2\sigma_0^2}[\sum_{i=1}^n x_i^2 + n - 2\sum_{i=1}^n x_i]}}{e^{\frac{-1}{2\sigma_0^2}\sum_{i=1}^n x_i^2}} = \frac{e^{\frac{-1}{2\sigma_0^2}\sum_{i=1}^n x_i^2} e^{\frac{-n}{2\sigma_0^2}} e^{\frac{1}{\sigma_0^2}\sum_{i=1}^n x_i}}{e^{\frac{-1}{2\sigma_0^2}\sum_{i=1}^n x_i^2}} > \frac{1}{n}. \end{aligned}$$

Tomando logaritmos,

$$\frac{-n}{2\sigma_0^2} + \frac{\sum x_i}{\sigma_0^2} > k' \iff \frac{\sum x_i}{\sigma_0^2} > k'' \iff \bar{x} > k'''$$

Determinamos k''' cuando $\mu = 0$ con la condición de que $\alpha = p_{H_0}(C)$. Es decir, $\alpha = p_{H_0}(C) = p_{\mu=0}(\bar{X} > k''')$. Como $\bar{X} \sim N\left(0, \frac{\sigma_0}{\sqrt{n}}\right)$, $p_{\mu=0}(\bar{X} > k''') = p\left(N\left(0, \frac{\sigma_0}{\sqrt{n}}\right) > k'''\right) = p\left(Z > \frac{k'''}{\frac{\sigma_0}{\sqrt{n}}}\sqrt{n}\right) \implies \frac{k'''}{\frac{\sigma_0}{\sqrt{n}}}\sqrt{n} = z_\alpha \implies k''' = \frac{z_\alpha \sigma_0}{\sqrt{n}}$, y la región crítica óptima de nivel α según Neyman-Pearson es $C = \{\bar{X} > \frac{z_\alpha \sigma_0}{\sqrt{n}}\}$.

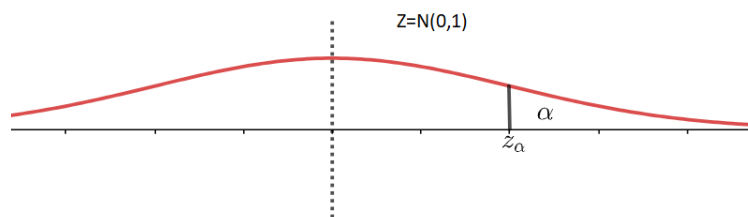
Por ejemplo, si $\alpha = 0,05$, $\sigma_0 = 0,25$ y $n = 100$, entonces $C = \{\bar{X} > \frac{1,645 \times 0,25}{\sqrt{100}}\} = \{\bar{X} > 0,041\}$.

Ejemplo 10.5:

Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim B(1, p)$. Se desea contrastar

$$H_0 : p = 0,02 \quad \text{frente a} \quad H_1 : p = 0,4$$

Obtener, aplicando el lema de Neyman-Pearson, la región crítica óptima de nivel α_0 .



$$C = \{x_1, x_2, \dots, x_n \mid f(x_1, x_2, \dots, x_n, 0, 2) < k f(x_1, x_2, \dots, x_n, 0, 4)\}$$

con k determinada para que $p_{H_0}(C) = \alpha$.

$$\frac{f(x_1, x_2, \dots, x_n, 0, 4)}{f(x_1, x_2, \dots, x_n, 0, 2)} = \frac{\prod_{i=1}^n 0,4^{x_i} \times 0,6^{1-x_i}}{\prod_{i=1}^n 0,2^{x_i} \times 0,8^{1-x_i}} = \frac{0,4^{\sum x_i} \times 0,6^{n-\sum x_i}}{0,2^{\sum x_i} \times 0,8^{n-\sum x_i}} = 2^{\sum x_i} \left(\frac{3}{4}\right)^{n-\sum x_i} = 2^{\sum x_i} \left(\frac{3}{4}\right)^n \left(\frac{4}{3}\right)^{\sum x_i} = \left(\frac{8}{3}\right)^{\sum x_i} \left(\frac{3}{4}\right)^n$$

Aplicando logaritmos,

$\left(\frac{8}{3}\right)^{\sum x_i} \left(\frac{3}{4}\right)^n > \frac{1}{k} \iff (\sum x_i) \left(\ln \frac{8}{3}\right) + n \ln \frac{3}{4} > \ln \frac{1}{k} \iff \sum x_i > \frac{\ln \frac{1}{k} - n \ln \frac{3}{4}}{\ln \frac{8}{3}} = k' \implies C = \{\sum_{i=1}^n X_i > k'\}$. Es decir, se rechaza cuando la frecuencia absoluta de éxitos es lo suficientemente grande, lo cual es intuitivo pues bajo H_1 , la probabilidad de éxito es mayor que bajo H_0 .

Sin embargo, al tratarse de una distribución discreta, no todos los niveles de α_0 podrán ser alcanzados.

Por ejemplo, siendo $n = 10$ y $\alpha_0 = 0,05$,

si $C = \{\sum_{i=1}^{10} X_i > 4\} = \{\sum_{i=1}^{10} X_i \geq 5\}$, entonces $\alpha_0 = 0,0328$;

si $C = \{\sum_{i=1}^{10} X_i > 3\} = \{\sum_{i=1}^{10} X_i \geq 4\}$, entonces $\alpha_0 = 0,1209$,

y por tanto, $\alpha_0 = 0,05$ no puede alcanzarse.

Para el caso en el que se disponga de muchas observaciones, $n \geq 30$, podemos utilizar el teorema central del límite, y obtener regiones críticas aproximadas de cualquier nivel.

Ejemplo 10.6:

Si $n = 100$, $\sum_{i=1}^{100} X_i \sim \text{Bin}(100, 0,02) \approx N(20, \sqrt{16}) = N(20, 4)$; entonces una región crítica óptima aproximada de nivel 0,05 será:

$$C = \{\sum X_i > k'\}, \text{ con } p_{H_0}(C) = 0,05 \iff$$

$$0,05 = p(\sum X_i > k') = 1 - p(\sum X_i \leq k') = 1 - p(N \leq k' + 0,5) = p(N(20, 4) > k') =$$

$$p\left(Z > \frac{k' + 0,5 - 20}{4}\right) \Rightarrow \frac{k' + 0,5 - 20}{4} = 1,645 \Rightarrow k' = 19,5 + 4 \times 1,645 = 19,5 + 6,58 = 26,08$$

y

$$C = \{\sum X_i > 26,08\} = \{\bar{X} > 0,2608\}$$

Recuérdese que el añadir 0.5 en el cálculo se corresponde con la corrección por continuidad para aproximar la binomial a la normal.

Sin embargo, existe una manera de conseguir, para tamaños muestrales pequeños, regiones críticas de nivel prefijado cualquiera. Consiste en recurrir a las llamadas **regiones críticas aleatorizadas**. En una región crítica aleatorizada, rechazaremos para ciertos valores, para otros mantendremos H_0 , y para un cierto valor rechazaremos sólo a veces, es decir, con una cierta probabilidad. Utilizaremos la función de probabilidad de la $Bin(10, 0,2)$.

| Bin(10, 0,2) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---------------------|-------|--------|--------|--------|--------|--------|--------|--------|--------|---|----|
| | 0.107 | 0.2684 | 0.3220 | 0.2015 | 0.0881 | 0.2640 | 0.0055 | 0.0088 | 0.0001 | 0 | 0 |

Tabla 3.1

Volviendo al ejemplo anterior, con $n = 10$ y $\alpha_0 = 0,05$, supongamos que utilizamos la siguiente región crítica

$$C^* = \begin{cases} \sum_{i=1}^{10} X_i \geq 5 & \text{rechazo } H_0 \\ \sum_{i=1}^{10} X_i = 4 & \text{rechazo } H_0 \text{ con probabilidad } p^* \text{ y acepto con probabilidad } 1 - p^* \\ \sum_{i=1}^{10} X_i < 4 & \text{mantengo } H_0 \end{cases}$$

Entonces, $p_{H_0}(C^*) = p_{H_0}(\sum X_i \geq 5) + p^* p(\sum X_i = 4) = 0,00325 + p^* 0,0771$. Igualando a 0,05, obtenemos el p^* que garantiza nivel 0,05:

$$0,0328 + p^* 0,0881 = 0,05 \implies p^* = \frac{0,0172}{0,0881} = \frac{172}{881}$$

Ejemplo 10.7:

Sea X_1, X_2, \dots, X_n una m.a.s. de $P(\lambda)$. Se desea contrastar:

$$H_0 : \lambda = 1 \quad \text{frente a } H_1 : \lambda = 3$$

Obtener una región crítica óptima aleatorizada de nivel $\alpha = 0,05$ para $n = 9$ observaciones. Además, obtener una región crítica óptima aproximada de nivel $\alpha = 0,05$ si $n = 64$ utilizando para ello el teorema central del límite.

Según el lema de Neyman-Pearson, y teniendo en cuenta que a partir de este punto se utilizará de forma permanente la notación muestral en lugar de la realización muestral, la región crítica óptima de nivel α es

$$C = \{X_1, \dots, X_n \mid \frac{f(X_1, \dots, X_n, 3)}{f(X_1, \dots, X_n, 1)} > k\}$$

determinándose k para que $p_{H_0}(C) = \alpha$.

$$\frac{f(X_1, \dots, X_n, 3)}{f(X_1, \dots, X_n, 1)} = \frac{\prod_{i=1}^n f(X_i, 3)}{\prod_{i=1}^n f(X_i, 1)} = \frac{\prod_{i=1}^n \left(e^{-3} \frac{3^{X_i}}{X_i!} \right)}{\prod_{i=1}^n \left(e^{-1} \frac{1^{X_i}}{X_i!} \right)} = \frac{e^{-3n} 3^{\sum X_i}}{e^{-n}} = e^{-2n} 3^{\sum X_i} > k$$

$$\underbrace{\Longleftrightarrow}_{\ln} -2n + \sum X_i \ln 3 > \underbrace{\ln k}_{k'} \Longleftrightarrow \sum X_i \ln 3 > \underbrace{k' + 2n}_{k''} \Longleftrightarrow \sum X_i > \underbrace{\frac{k''}{3}}_{k'''}$$

$$C = \{X_1, \dots, X_n \mid \sum X_i > k'''\}$$

Si $n = 9$, la distribución de $\sum_{i=1}^9 X_i$ bajo H_0 es *Poisson*(9), ya que es la suma de 9 variables aleatorias independientes de *Poisson*(1).

La constante k''' se determina para que $0,05 = p_{H_0}(C)$. Ahora bien, por ser discreta la *Poisson*(9), no necesariamente podemos garantizar que se consiga el nivel $\alpha = 0,05$.

| <i>Poisson</i> (9) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|--------------------|--------|-------|--------|--------|--------|--------|--------|--------|--------|------|--------|--------|--------|--------|--------|--------|
| | 0.0001 | 0.007 | 0.0034 | 0.0107 | 0.0254 | 0.0483 | 0.0764 | 0.1037 | 0.1232 | 0.13 | 0.1235 | 0.1067 | 0.0844 | 0.0617 | 0.0419 | 0.0265 |

Si $C = \{\sum_{i=1}^9 X_i \geq 16\}$, entonces $p_{H_0}(C) = 0,0187 + 0,0088 + 0,0046 + 0,0023 + 0,0011 + 0,0005 + 0,0002 + 0,0001 = 0,0363$

Si $C = \{\sum_{i=1}^9 X_i \geq 15\}$, entonces $p_{H_0}(C) = 0,0363 + 0,0265 = 0,0628 > 0,05$

Podemos obtener una regla de decisión con región crítica de nivel $\alpha = 0,05$ aleatorizado.

- Si $\sum_{i=1}^9 X_i \geq 16$, rechazamos H_0 .
- Si $\sum_{i=1}^9 X_i = 15$, rechazamos H_0 con p^* .
- Si $\sum_{i=1}^9 X_i < 15$, mantenemos H_0 .

Determinación de p^*

$$\begin{aligned}\alpha = 0,05 &= p(\text{rechazar } H_0 \text{ si es cierta}) = p_{H_0}(\sum X_i \geq 16) + p_{H_0}(\sum X_i = 15) p^* = \\ &= 0,0363 + p^* 0,0265 = 0,05 \implies p^* = \frac{137}{265}\end{aligned}$$

Si $n = 64$, se puede utilizar el teorema central del límite, de manera que

$$C = \{X_1, \dots, X_n \mid \sum_{i=1}^{64} X_i \geq k'''\} = \{X_1, \dots, X_n \mid \bar{X} \geq k^*\}$$

Bajo $H_0 : \lambda = 1$, por el teorema central del límite, $\bar{X} \rightsquigarrow N(1, \frac{1}{8})$, con lo que $k^* = \frac{k'''}{64}$.

En este caso, $\alpha = 0,05 = p_{H_0}(\bar{X} > k^*) = p_{H_0}\left(N(0, 1) \geq \frac{k^* + 0,5 - 1}{\frac{1}{8}}\right)$; $\frac{k^* + 0,5 - 1}{\frac{1}{8}} = 1,645 \implies k^* = 0,5 + \frac{1,645}{8} \approx 1,706$.

Ejemplo 10.8:

Sea X_1, X_2, \dots, X_n una m.a.s de $X \rightsquigarrow \exp(\lambda)$. Se desea contrastar

$$H_0 : \lambda = \frac{1}{2} \quad \text{frente a} \quad H_1 : \lambda = 1$$

con n observaciones.

Obtener la región crítica óptima de nivel α . Particularizar el caso de $\alpha = 0,05$ y $n = 15$.

En este caso, por ser la distribución continua, podemos obtener una región crítica de nivel exactamente α .

Según el teorema de Neyman-Pearson, tal región es

$$C = \left\{X_1, \dots, X_n \mid \frac{f(X_1, \dots, X_n, 1)}{f(X_1, \dots, X_n, \frac{1}{2})} > k\right\}$$

$$\frac{f(X_1, \dots, X_n, 1)}{f(X_1, \dots, X_n, \frac{1}{2})} = \frac{\prod f(X_i, 1)}{\prod f(X_i, \frac{1}{2})} = \frac{\prod_{i=1}^n (1e^{-X_i})}{\prod_{i=1}^n \left(\frac{1}{2}e^{-\frac{1}{2}X_i}\right)} = \frac{e^{-\sum_{i=1}^n X_i}}{\left(\frac{1}{2}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n X_i}} = 2^n e^{\frac{-1}{2}\sum_{i=1}^n X_i} > k$$

$$\iff n \ln 2 - \frac{1}{2} \sum X_i > \underbrace{\ln k}_{k'} \iff \frac{-1}{2} \sum X_i > k' - n \ln 2 \iff \underbrace{\sum X_i < -2k' + 2n \ln 2}_{k''}$$

$$\implies C = \{X_1, \dots, X_n \mid \sum_{i=1}^n X_i < k''\}$$

y k'' se obtiene de $\alpha = p_{H_0}(C)$.

Ahora bien, si $\sum X_i$ es la suma de variables independientes, $\exp(\frac{1}{2}) = \Gamma(1, \frac{1}{2})$ bajo H_0 , por la reproductividad de la distribución gamma, $\sum X_i \sim \Gamma(n, \frac{1}{2}) \equiv \Gamma(\frac{2n}{2}, \frac{1}{2}) \equiv \chi_{2n}^2$, luego $\alpha = p_{H_0}(\sum X_i \geq k'') \implies k'' = \chi_{2n, \alpha}^2$.

En particular, si $\alpha = 0,05$ y $n = 15$, $\chi_{30, 0,05}^2 = 43,773$. Si la suma de las 15 observaciones supera el valor 43,773, se rechaza H_0 .

10.5 Contrastes de hipótesis unilaterales

Sea X_1, \dots, X_n una m.a.s. de $X \sim f(x, \theta)$. Sea θ_0 un valor fijo del parámetro. Diremos que un test es de hipótesis unilaterales si ambas, la nula y la alternativa, lo son:

$$\begin{aligned} H_0 : \theta < \theta_0 & \text{ frente a } H_1 : \theta > \theta_0 \\ H_0 : \theta > \theta_0 & \text{ frente a } H_1 : \theta < \theta_0 \end{aligned}$$

Adoptaremos la primera opción, pero la segunda es muy similar.

Definición 10.2:

Dada una muestra aleatoria simple X_1, X_2, \dots, X_n de $X \sim f(x, \theta)$, diremos que la distribución tiene cociente monótono en el estadístico T si para $\theta \leq \theta'$, $\frac{f(X_1, \dots, X_n, \theta')}{f(X_1, \dots, X_n, \theta)}$ es creciente en T .

Ejemplo 10.9:

Sea X_1, \dots, X_n una m.a.s. de $X \sim B(1, p)$, y sean $p < p'$. Entonces,

$$\begin{aligned} \frac{f(X_1, \dots, X_n, p')}{f(X_1, \dots, X_n, p)} &= \frac{\prod_{i=1}^n f(X_i, p')}{\prod_{i=1}^n f(X_i, p)} = \frac{\prod (p'^{X_i} (1-p')^{1-X_i})}{\prod (p^{X_i} (1-p)^{1-X_i})} = \\ &= \frac{p'^{\sum X_i} (1-p')^{n-\sum X_i}}{p^{\sum X_i} (1-p)^{n-\sum X_i}} = \left(\frac{p'}{p}\right)^{\sum X_i} \left(\frac{1-p'}{1-p}\right)^{n-\sum X_i} \end{aligned}$$

Entonces, la distribución $B(1, p)$ tiene cociente monótono en $T = \sum X_i$.

Ejemplo 10.10:

Sea X_1, X_2, \dots, X_n una m.a.s. de $\text{Bin}(m, p)$. Determinar un estadístico en el que el cociente de verosimilitudes sea monótono. Siendo $p < p'$,

$$\frac{f(X_1, \dots, X_n, p')}{f(X_1, \dots, X_n, p)} = \frac{\prod_{i=1}^n f(X_i, p')}{\prod_{i=1}^n f(X_i, p)} = \frac{\prod_{i=1}^n \left[\binom{m}{X_i} p'^{\Sigma X_i} (1-p')^{mn-\Sigma X_i} \right]}{\prod_{i=1}^n \left[\binom{m}{X_i} p^{\Sigma X_i} (1-p)^{mn-\Sigma X_i} \right]} =$$

$$\left(\frac{p'}{p} \right)^{\Sigma X_i} \left(\frac{1-p'}{1-p} \right)^{mn-\Sigma X_i} = \left(\frac{p'}{p} \right)^{\Sigma X_i} \left(\frac{1-p'}{1-p} \right)^{\Sigma X_i - mn}$$

La distribución de $\text{Bin}(m, p)$ tiene cociente monótono en $T = \Sigma X_i$.

Ejemplo 10.11:

Sea X_1, X_2, \dots, X_n una m.a.s. de $\text{Poisson}(\lambda)$. Determinar un estadístico en el que el cociente de verosimilitudes sea monótono.

Si $\lambda < \lambda'$,

$$\frac{f(X_1, \dots, X_n, \lambda')}{f(X_1, \dots, X_n, \lambda)} = \frac{\prod_{i=1}^n \left[\frac{e^{-\lambda'} \lambda'^{X_i}}{X_i!} \right]}{\prod_{i=1}^n \left[\frac{e^{-\lambda} \lambda^{X_i}}{X_i!} \right]} = \frac{e^{-n\lambda'} \left[\prod_{i=1}^n \frac{1}{X_i!} \right] \lambda'^{\Sigma X_i}}{e^{-n\lambda} \left[\prod_{i=1}^n \frac{1}{X_i!} \right] \lambda^{\Sigma X_i}} = \underbrace{e^{-n\lambda' + n\lambda}}_{\text{constante positiva}} \left(\frac{\lambda'}{\lambda} \right)^{\Sigma X_i}$$

La distribución de $\text{Poisson}(\lambda)$ tiene cociente monótono en $T = \Sigma X_i$.

Ejemplo 10.12:

Sea X_1, \dots, X_n una m.a.s. de $\Gamma(a, p_0)$ con p_0 conocido y a desconocido. Determinar un estadístico en el que el cociente de verosimilitudes sea monótono.

Si $a < a'$,

$$\frac{f(X_1, \dots, X_n, a')}{f(X_1, \dots, X_n, a)} = \frac{\prod_{i=1}^n f(X_i, a')}{\prod_{i=1}^n f(X_i, a)} = \frac{\prod_{i=1}^n \left[\frac{p_0^{a'}}{\Gamma(a')} e^{-p_0 X_i} X_i^{a'-1} \right]}{\prod_{i=1}^n \left[\frac{p_0^a}{\Gamma(a)} e^{-p_0 X_i} X_i^{a-1} \right]} = \frac{\left(\frac{p_0^{a'}}{\Gamma(a')} \right)^n e^{-p_0 \Sigma X_i} (\prod X_i)^{a'-1}}{\left(\frac{p_0^a}{\Gamma(a)} \right)^n e^{-p_0 \Sigma X_i} (\prod X_i)^{a-1}} =$$

$$\frac{p_0^{n(a'-a)}}{\frac{\Gamma(a')}{\Gamma(a)}} \left(\prod_{i=1}^n X_i \right)^{a'-a}$$

y para $T = \prod_{i=1}^n X_i$ el cociente de verosimilitudes es monótono.

Ejemplo 10.13:

Sea X_1, X_2, \dots, X_n una m.a.s de $\Gamma(a, p)$, con a conocido y $p < p'$ desconocidos.

$$\frac{f(X_1, \dots, X_n, p')}{f(X_1, \dots, X_n, p)} = \frac{\frac{p'^{an}}{[\Gamma(a)]^n} e^{-p' \sum X_i} \prod X_i^{a-1}}{\frac{p^{an}}{[\Gamma(a)]^n} e^{-p \sum X_i} \prod X_i^{a-1}} = \left(\frac{p'}{p}\right)^{an} e^{-p' \sum X_i + p \sum X_i} = \left(\frac{p'}{p}\right)^{an} e^{\sum X_i (p-p')} \Rightarrow$$

$T = -\sum X_i$ es un estadístico en el que el cociente de verosimilitud es monótono creciente.

El caso particular de $X \sim \Gamma(1, \lambda) = \exp(\lambda)$ nos lleva a $T = -\sum_{i=1}^n X_i$ como estadístico en el que el cociente es monótono.

Si $\lambda < \lambda'$,

$$\frac{f(X_1, \dots, X_n, \lambda')}{f(X_1, \dots, X_n, \lambda)} = \frac{\prod_{i=1}^n (\lambda' e^{-\lambda' X_i})}{\prod_{i=1}^n (\lambda e^{-\lambda X_i})} = \left(\frac{\lambda'}{\lambda}\right)^n e^{-\sum X_i (\lambda - \lambda')}$$

Ejemplo 10.14:

Sea X_1, X_2, \dots, X_n una m.a.s. de $N(\mu, \sigma_0)$. Encontrar un estadístico en el que el cociente sea monótono, con σ_0 conocida y $\mu < \mu'$ desconocidas.

$$\begin{aligned} \frac{f(X_1, \dots, X_n, \mu')}{f(X_1, \dots, X_n, \mu)} &= \frac{\prod_{i=1}^n f(X_i, \mu')}{\prod_{i=1}^n f(X_i, \mu)} = \frac{\prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}\sigma_0} \exp \left[\frac{-1}{2\pi\sigma_0^2} (X_i - \mu')^2 \right] \right]}{\prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}\sigma_0} \exp \left[\frac{-1}{2\pi\sigma_0^2} (X_i - \mu)^2 \right] \right]} = \\ &= \frac{\exp \left(\frac{-1}{2\sigma_0^2} \sum X_i^2 \right) \exp \left(\frac{-1}{2\sigma_0^2} n\mu'^2 \right) \exp \left(\frac{\sum X_i \mu'}{\sigma_0^2} \right)}{\exp \left(\frac{-1}{2\sigma_0^2} \sum X_i^2 \right) \exp \left(\frac{-1}{2\sigma_0^2} n\mu^2 \right) \exp \left(\frac{\sum X_i \mu}{\sigma_0^2} \right)} = \exp \left[\frac{-1}{2\sigma_0^2} n(\mu'^2 - \mu^2) \right] \exp \left[\frac{\sum X_i}{\sigma_0^2} (\mu' - \mu) \right] \end{aligned}$$

Luego $T = \sum X_i$ es un estadístico en el que el cociente es monótono.

Ejemplo 10.15:

Sea X_1, X_2, \dots, X_n una m.a.s. de $N(\mu_0, \sigma)$. Encontrar un estadístico en el que el cociente sea monótono, con μ conocida y $\sigma < \sigma'$ desconocidas.

$$\frac{f(X_1, \dots, X_n, \sigma')}{f(X_1, \dots, X_n, \sigma)} = \frac{\prod_{i=1}^n f(X_i, \sigma')}{\prod_{i=1}^n f(X_i, \sigma)} = \frac{\prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}\sigma'} e^{\frac{-1}{2\sigma'^2} (X_i - \mu_0)^2} \right]}{\prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2} (X_i - \mu_0)^2} \right]} = \frac{\sigma^n}{\sigma'^n} e^{\left(\frac{-1}{2\sigma'^2} + \frac{1}{2\sigma^2} \right) \sum_{i=1}^n (X_i - \mu_0)^2}$$

Por tanto, $T = \sum_{i=1}^n (X_i - \mu_0)^2$ es un estadístico en el que el cociente es monótono creciente.

Una vez que trabajamos con hipótesis compuestas, el error de tipo I y el error de tipo II, especialmente sus probabilidades α y β respectivamente, no pueden ser calculadas. Todavía se puede hablar de error de tipo I como el error obtenido al rechazar H_0 siendo cierta, y error de tipo II como el error de mantener H_0 siendo falsa. Sin embargo, la probabilidad

$$\alpha = p_{H_0}(C) \quad \text{ya no está definida si } H_0 \text{ es compuesta}$$

$$\beta = p_{H_1}(\bar{C}) \quad \text{ya no está definida si } H_0 \text{ es compuesta}$$

La optimalidad de las reglas de decisión para hipótesis simples que se obtenía a partir del lema de Neyman-Pearson, pivotaba sobre los valores de α y β . Para contrastes de hipótesis compuestas, se requiere un criterio diferente. Definiremos, entonces, la función de potencia de un contraste.

Función de potencia:

Sea X_1, \dots, X_n una m.a.s. de $X \sim f(x, \theta)$, con θ desconocido, $\theta \in \Theta$ (espacio paramétrico). Supongamos que se desea llevar a cabo el contraste

$$H_0 : \theta \in \Theta_0 \quad \text{frente a} \quad H_1 : \theta \in \Theta_1 = \overline{\Theta_0}$$

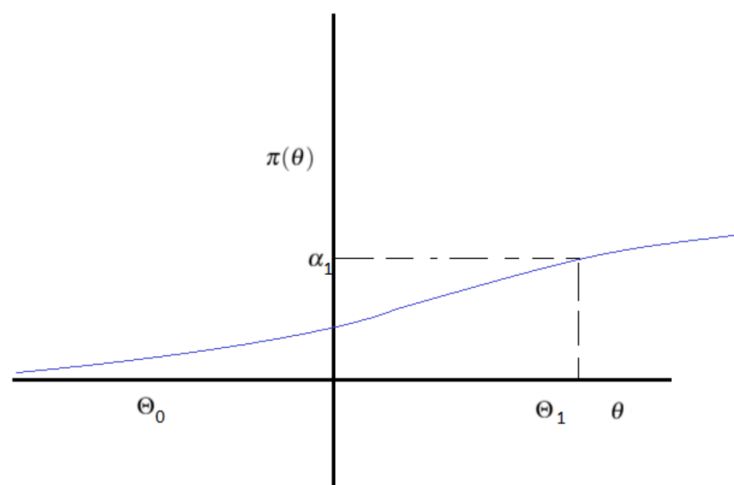
Llamaremos función de potencia del contraste a

$$\begin{aligned} \pi : \Theta &\longrightarrow [0, 1] \\ \theta &\longrightarrow \pi(\theta) = p_\theta(C), \end{aligned} \tag{10.3}$$

siendo C la región crítica del contraste.

La función de potencia evalúa, para cada valor del parámetro, la probabilidad de rechazar la hipótesis nula H_0 si éste es el verdadero valor del parámetro.

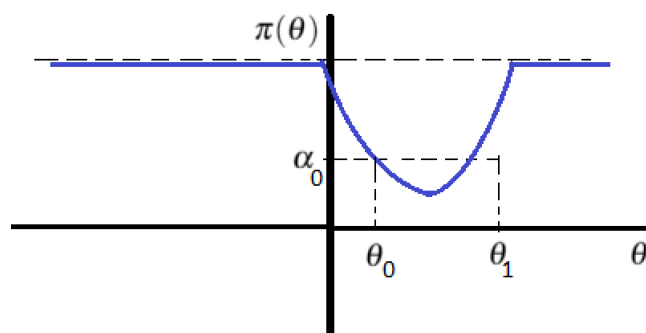
Una región crítica “apropiada” para el contraste $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$ tiene función de potencia del tipo:



Una región crítica “apropiada” para un contraste

$$H_0 : \theta \in [\theta_0, \theta_1] \quad \text{frente a} \quad H_1 : \theta < \theta_0 \quad \text{ó} \quad \theta > \theta_1$$

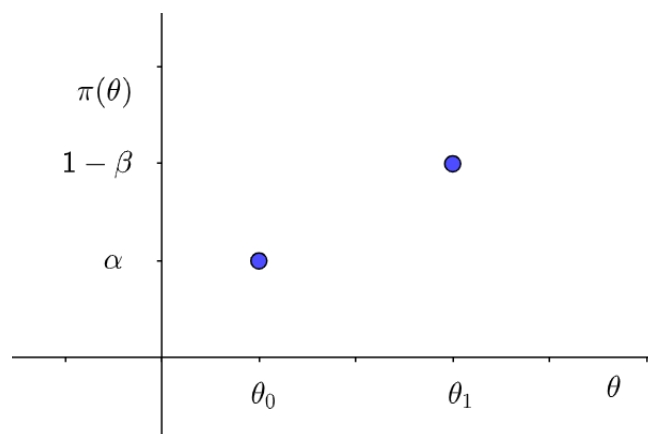
tendría una función de potencia del tipo:



Observación. Los casos de hipótesis simples analizados hasta ahora:

$$H_0 : \theta = \theta_0 \quad \text{frente a} \quad H_1 : \theta = \theta_1$$

tienen como función de potencia $\pi(\theta)$:



donde

$$\alpha = p_{H_0}(C)$$

$$\beta = p_{H_1}(\bar{C}) = 1 - p_{H_0}(C) = 1 - \pi(\theta)$$

Ejemplo 10.16:

Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim N(\mu, 1)$. Se desea contrastar

$$H_0 : \mu \leq 0 \quad \text{frente a} \quad H_1 : \mu > 0$$

Determinar la función de potencia asociada a la regla de decisión con región crítica $C = \{\bar{X} > 1,2\}$.

$$\pi(\mu) = p_\mu(C) = p_\mu(\bar{X} > 1,2) \underset{\bar{X} \sim N(\mu, \frac{1}{\sqrt{n}})}{=} p\left(N\left(\mu, \frac{1}{\sqrt{n}}\right) > 1,2\right) = p(Z > (1,2 - \mu)\sqrt{n})$$

Supongamos que $n = 25$. Entonces:

- $\pi(0) = p(Z > 6) = 0$
- $\pi(1) = p(Z > 1) = 0,1587$
- $\pi(3) = p(Z > -9) = 1$
- $\pi(-1) = p(Z > 11) = 0$.

Definición 10.3:

Para un contraste de hipótesis unilaterales, diremos que una región crítica C^* es uniformemente más potente de nivel de significación α si su función de potencia $\pi_{C^*}(\theta)$ satisface:

- (a) $\pi_{C^*}(\theta) \leq \alpha, \forall \theta \leq \theta_0$.
- (b) Para cada región crítica C , que cumple: $\pi_C(\theta) \leq \alpha$ si $\theta < \theta_0$, entonces $\pi_C(\theta) > \pi_{C^*}(\theta)$, si $\theta > \theta_0$.

Teorema 10.1

Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim f(x, \theta)$, siendo θ un parámetro desconocido. Se desea contrastar las siguientes hipótesis compuestas unilaterales:

$$H_0 : \theta \leq \theta_0 \quad \text{frente a} \quad H_1 : \theta > \theta_0$$

siendo θ_0 un valor fijo del parámetro.

La región crítica uniformemente más potente (U.M.P) de nivel de significación α para este contraste es:

$$C^* = \{X_1, \dots, X_n \mid T \geq K\}$$

siendo T el estadístico en el que el cociente de verosimilitudes sea monótono para $f(x, \theta)$. El valor de la constante K se calcula con la condición $p_{\theta_0}(C^*) = \alpha$.

Demostración:

Por un lado, $p_\theta(C^*)$ es creciente con θ por ser el cociente de verosimilitudes monótono (creciente) en θ . Si fijamos K para que $p_{\theta_0}(C^*) = \alpha$, entonces $p_\theta(C^*) \leq p_{\theta_0}(C^*) = \alpha$, y por tanto, se satisface la condición (a) de la definición de región crítica U.M.P.

Por otro lado, sea C una región crítica tal que $p_\theta(C) \leq \alpha, \forall \theta \leq \theta_0$. Por el lema de Neyman-Pearson, sabemos que $\forall \theta > \theta_0, p_\theta(C) \leq p_\theta(C^*)$.

Intuitivamente, C^* es óptima, porque al ser el cociente de verosimilitudes monótono en T , cuando T crece, crece la verosimilitud de la muestra bajo valores grandes de θ . Por tanto, valores grandes de T apoyan más la hipótesis alternativa. Por eso, estos valores pertenecen a la región crítica.

Corolario

Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim f(x, \theta)$, siendo θ un parámetro desconocido. Si T es el estadístico en el que el cociente de verosimilitudes es monótono, para contrastar

$$H_0 : \theta \geq \theta_0 \quad \text{frente a} \quad H_1 : \theta < \theta_0$$

la región crítica U.M.P. de nivel α es

$$C^* = \{X_1, \dots, X_n \mid T \leq K\}$$

siendo K el valor que verifica $p_{\theta_0}(C^*) = \alpha$.

Ejemplo 10.17:

Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim U(0, \theta)$. Se desea contrastar las hipótesis

$$H_0 : \theta \leq 1 \quad \text{frente a} \quad H_1 : \theta > 1$$

Determinar si existe una región crítica U.M.P. de nivel α , y calcúlese en dicho caso.

Existirá tal región si la distribución $U(0, \theta)$ tiene cociente monótono en algún estadístico T . Supongamos que $\theta < \theta'$

$$\frac{f(X_1, \dots, X_n, \theta')}{f(X_1, \dots, X_n, \theta)} = \frac{\prod_{i=1}^n f(X_i, \theta')}{\prod_{i=1}^n f(X_i, \theta)} = \frac{\prod_{i=1}^n \frac{1}{\theta'}}{\prod_{i=1}^n \frac{1}{\theta}} = \frac{\left(\frac{1}{\theta'}\right)^n}{\left(\frac{1}{\theta}\right)^n} \cdot \frac{h_1(X_1, \dots, X_n)}{h_2(X_1, \dots, X_n)} \cdot \frac{g_1(X_1, \dots, X_n, \theta')}{g_2(X_1, \dots, X_n, \theta)}$$

$$\frac{\left(\frac{1}{\theta'}\right)^n}{\left(\frac{1}{\theta}\right)^n} \cdot \frac{h_1(X_1, \dots, X_n)}{h_2(X_1, \dots, X_n)} \cdot \frac{g_1(X_1, \dots, X_n, \theta')}{g_2(X_1, \dots, X_n, \theta)} = \frac{\theta^n}{(\theta')^n} \cdot \frac{h_1(X_1, \dots, X_n)}{h_2(X_1, \dots, X_n)} \cdot \frac{g_1(X_1, \dots, X_n, \theta')}{g_2(X_1, \dots, X_n, \theta)}$$

con

$$h_1(X_1, \dots, X_n) = \begin{cases} 1 & \text{si } \min\{X_1, \dots, X_n\} \geq 0 \\ 0 & \text{en otro caso,} \end{cases}$$

siendo

$$g_1(X_1, \dots, X_n, \theta') = \begin{cases} 1 & \text{si } \max\{X_1, \dots, X_n\} \leq \theta' \\ 0 & \text{en otro caso} \end{cases}$$

$$g_2(X_1, \dots, X_n, \theta) = \begin{cases} 1 & \text{si } \max\{X_1, \dots, X_n\} \leq \theta \\ 0 & \text{en otro caso,} \end{cases}$$

luego

$$\frac{f(X_1, \dots, X_n, \theta')}{f(X_1, \dots, X_n, \theta)} = \frac{\theta^n}{(\theta')^n} \cdot \frac{g_1(X_1, \dots, X_n, \theta')}{g_2(X_1, \dots, X_n, \theta)}.$$

Justificamos que $T = \max\{X_1, \dots, X_n\}$ es el estadístico en el que el cociente es monótono. Cuando T crece, pueden pasar dos cosas:

(a) $g_1 = g_2 = 1$ si $T < \theta$ y $T < \theta'$.

(b) $g_1 = 1$ y $g_2 = 0$ si $T < \theta'$ pero $T > \theta$.

Lo que nunca puede ocurrir es que $g_1 = 0$ y $g_2 = 1$, pues entonces $\max\{X_1, \dots, X_n\} \leq \theta$ y $\max\{X_1, \dots, X_n\} \geq \theta$ y $\theta' > \theta$, lo que nos lleva a un absurdo.

En cualquier caso, al crecer T se incrementa el valor de g_1 , o en el peor de los casos se mantiene, frente al valor de g_2 ; esto es, la verosimilitud crece con T .

Por tanto, existe región crítica U.M.P. de nivel α para este contraste, y es:

$$C^* = \{X_1, \dots, X_n \mid \max\{X_1, \dots, X_n\} \geq k\}$$

con la condición $p_{\theta=1}(C^*) = \alpha$, con

$$\alpha = p_{\theta=1}(C^*) = p_{\theta=1}(\max\{X_1, \dots, X_n\} \geq K).$$

Recordemos que

$$f_T(t) = \begin{cases} \frac{nt^{n-1}}{\theta^n} & \text{si } t \in (0, \theta) \\ 0 & \text{en otro caso,} \end{cases}$$

luego

$$p_{\theta=1}(C^*) = \int_K^1 nt^{n-1} dt = [t^n]_K^1 = 1 - K^n \implies K^n = 1 - \alpha \implies K = \sqrt[n]{1 - \alpha}.$$

Por ejemplo, si $\alpha = 0,05$ y $n = 25$, la región crítica U.M.P. para el contraste $H_0 : \theta \leq 1$ frente a $\theta > 1$ es $C^* = \{X_1, \dots, X_{25} \mid \max\{X_1, \dots, X_{25}\} \geq 0,9979\}$. Si el máximo de las 25 observaciones supera el valor 0,9979, no podremos mantener H_0 .

10.6 Test de hipótesis compuestas (no unilaterales).

Sea X_1, \dots, X_n una m.a.s. de $X \rightsquigarrow f(x, \theta)$. Se desea contrastar

$$H_0 : \theta \in \Theta_0 \quad \text{frente a} \quad H_1 : \theta \in \Theta_1$$

siendo Θ_0 y Θ_1 conjuntos disjuntos del espacio paramétrico.

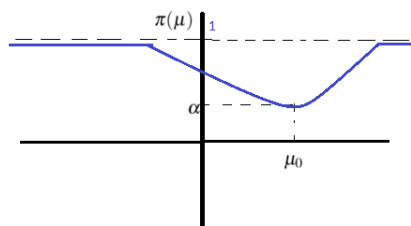
En este contexto ya no es posible obtener una región crítica óptima U.M.P. en el sentido de que si C^* es tal región y $\pi_{C^*}(\theta)$ es función de potencia, entonces

$$\begin{aligned} \pi_{C^*}(\theta) &\leq \alpha, \forall \theta \in \Theta_0 \\ \pi_{C^*}(\theta) &\geq \pi_C(\theta), \forall \theta \in \Theta_1 \end{aligned}$$

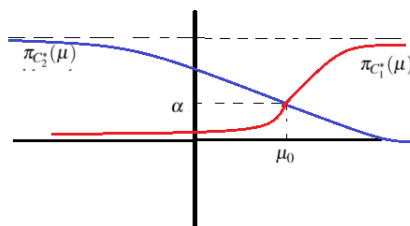
siendo C cualquier otra región crítica.

Ejemplo 10.18:

Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim N(\mu, 1)$. Se desea contrastar $H_0 : \mu = \mu_0$ frente a $H_1 : \mu \neq \mu_0$ (hipótesis nula simple frente a hipótesis alternativa bilateral). La región crítica U.M.P. de nivel α , si existiera para este contraste, sería tal que su función de potencia verificaría:



Notemos que para contrastar $H_0 : \mu \leq \mu_0$ frente a $H_1 : \mu > \mu_0$, por ser hipótesis unilaterales, existiría una región crítica U.M.P. de nivel α , C_1^* , que verificaría $\pi_{C_1^*}(\mu) \leq \alpha, \forall \mu \leq \mu_0$ y $\pi_{C_1^*}(\mu)$ máxima $\forall \mu > \mu_0$.



Así mismo, para contrastar $H_0 : \mu > \mu_0$ frente a $H_1 : \mu \leq \mu_0$, también existe una región crítica U.M.P. de nivel α por ser hipótesis unilaterales. Llamaremos C_2^* a la región que verifica $\pi_{C_2^*}(\mu) \leq \alpha, \forall \mu > \mu_0$ y $\pi_{C_2^*}(\mu)$ máxima $\forall \mu \leq \mu_0$.

Entonces, no es posible obtener una región crítica C en el contraste $H_0 : \mu = \mu_0$ frente a $H_1 : \mu \neq \mu_0$ tal que $\pi_C(\mu)$ será máxima si $\mu \neq \mu_0$. La razón es que:

$$\forall \mu > \mu_0, \pi_{C_1^*}(\mu) \geq \pi_C(\mu)$$

$$\forall \mu \leq \mu_0, \pi_{C_2^*}(\mu) \geq \pi_C(\mu)$$

verificando ambas que $\pi_{C_1^*}(\mu) = \pi_{C_2^*}(\mu) = \alpha$.

10.6.1 Test de la razón de verosimilitudes

Sea X_1, \dots, X_n una m.a.s. de $X \sim f(x, \theta)$. Se desea contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$. A pesar de no existir una región crítica U.M.P., se mantiene la idea natural de rechazar H_0 cuando la verosimilitud bajo H_1 sea “grande” respecto a la verosimilitud bajo H_0 . Esto da lugar

al llamado **test de la razón de verosimilitudes**, que se define de la siguiente manera: su región crítica es

$$C = \{X_1, \dots, X_n \mid \frac{\sup_{\theta \in \Theta_1} f(X_1, \dots, X_n, \theta)}{\sup_{\theta \in \Theta_0} f(X_1, \dots, X_n, \theta)} \geq K\}$$

determinándose K para que $\pi_C(\theta) \leq \alpha$ si $\theta \in \Theta_0$, con nivel de significación prefijado, α .

10.6.2 Test de razón de verosimilitudes en caso de normalidad

Test Z sobre la media Sea X_1, \dots, X_n una m.a.s. de $N(\mu, \sigma_0)$. Se desea contrastar $H_0 : \mu = \mu_0$ frente a $H_1 : \mu \neq \mu_0$ con nivel α . La región crítica proporcionada por el test de razón de verosimilitudes es:

$$C = \{X_1, \dots, X_n \mid \left| \frac{\bar{X} - \mu_0}{\sigma_0} \right| > z_{\frac{\alpha}{2}}\} = \left\{ \begin{array}{l} \bar{X} > \mu_0 + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}, \quad \text{ó} \\ \bar{X} < \mu_0 - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}, \end{array} \right.$$

Ejemplo 10.19:

En el control de calidad de una fábrica de latas de sardinas se desea analizar si el peso medio es de 1000 gramos, considerando para ello 100 latas de sardinas. Asumiendo $\sigma_0 = 5gr$.

$$H_0 : \mu = 1000gr \quad \text{frente a} \quad H_1 : \mu \neq 1000gr$$

La región crítica del test de razón de verosimilitudes de nivel 0,05 es:

$$C = \left\{ \left| \frac{\bar{X} - 1000}{5} \sqrt{100} \right| > 1,96 \right\} = \left\{ \bar{X} > 1000 + 0,98 = 1000,98 \quad \text{ó} \quad \bar{X} < 1000 - 0,98 = 999,02 \right\}$$

Rechazará el peso promedio de 1000gr si la media de los pesos de las latas (100 latas) supera 1000,98, o se queda por debajo de 999,02.

Test t sobre la media

Sea X_1, \dots, X_n una m.a.s. de $N(\mu, \sigma)$, con μ, σ desconocidos. Se desea contrastar $H_0 : \mu = \mu_0$ frente a $H_1 : \mu \neq \mu_0$ con nivel α . La región crítica proporcionada por el test de razón de verosimilitudes es:

$$C = \left\{ X_1, \dots, X_n \mid \left| \frac{\bar{X} - \mu_0}{S} \sqrt{n-1} \right| > t_{n-1, \frac{\alpha}{2}} \right\} = \left\{ X_1, \dots, X_n \mid \begin{array}{l} \bar{X} > \mu_0 + \frac{S}{\sqrt{n-1}} t_{n-1, \frac{\alpha}{2}}, \quad \text{ó} \\ \bar{X} < \mu_0 - \frac{S}{\sqrt{n-1}} t_{n-1, \frac{\alpha}{2}}, \end{array} \right\}$$

La región crítica es el complementario del intervalo de confianza para μ de nivel $1 - \alpha$.

Ejemplo 10.20:

Si para el mismo contexto del ejemplo anterior se supone que la desviación típica observada en la muestra es $S = 5gr$, entonces

$$C = \left\{ \left| \frac{\bar{X} - 1000}{S} \sqrt{99} \right| > t_{99, \frac{\alpha}{2}} \right\} = \left\{ \bar{X} > 1000 + \frac{5}{\sqrt{99}} 1,99 \quad \text{ó} \quad \bar{X} < 1000 - \frac{5}{\sqrt{99}} 1,99 \right\}$$

Test χ^2 para la varianza de una normal

Sea X_1, \dots, X_n una m.a.s. de $N(\mu, \sigma)$, con μ y σ desconocidos. Se desea contrastar $H_0 : \sigma^2 = \sigma_0^2$ frente a $H_1 : \sigma^2 \neq \sigma_0^2$. Siendo α el nivel de significación prefijado, la región crítica para este contraste es:

$$C = \left\{ X_1, \dots, X_n \mid \frac{nS^2}{\sigma_0^2} > \chi_{n-1, \frac{\alpha}{2}}^2 \quad \text{ó} \quad \frac{nS^2}{\sigma_0^2} < \chi_{n-1, 1-\frac{\alpha}{2}}^2 \right\}$$

Test Z para la diferencia de medias en poblaciones normales.

Sea X_1, \dots, X_n una m.a.s. de $N(\mu_1, \sigma_1)$ y Y_1, \dots, Y_n una m.a.s. de $N(\mu_2, \sigma_2)$, ambas muestras independientes. Supondremos que μ_1 y μ_2 son parámetros desconocidos, mientras que σ_1 y σ_2 son conocidos. Se desea contrastar, con nivel de significación α , las siguientes hipótesis:

$$H_0 : \mu_1 - \mu_2 = K \quad \text{frente a} \quad H_1 : \mu_1 - \mu_2 \neq K$$

En general, el valor de K con mayor interés es $K = 0$, es decir, contrastar medias iguales frente a medias diferentes.

La región crítica que proporciona el test de razón de verosimilitudes para nivel α es:

$$C = \left\{ X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2} : \left| \frac{\bar{X} - \bar{Y} - K}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right| > z_{\frac{\alpha}{2}} \right\}$$

Ejemplo 10.21:

Se asume que la estatura de los jóvenes varones españoles de 20 años en el 1980 era $N(\mu_1, 12cm)$, y en el año 1990 era $N(\mu_2, 14cm)$. Elegidas dos muestras independientes de tamaño $n_1 = 100$ y $n_2 = 225$, se obtuvo que la estatura media de los 100 jóvenes de 1980 fue $172cm$, mientras que la de los 225 de 1990 fue $175cm$. Con estos datos, contrastar a nivel $\alpha = 0,05$ si se puede asumir la igualdad de las medias poblacionales.

$$H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$$

$$\frac{\bar{X} - \bar{Y} - K}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{172 - 175}{\sqrt{\frac{12^2}{100} + \frac{14^2}{225}}} = -1,97$$

$$\left| \frac{\bar{X} - \bar{Y} - K}{\sqrt{\frac{\sigma_1^2}{n_1} + \sigma_2^2 n_2}} \right| = 1,97 > 1,96 = z_{0,025} \implies \text{se rechaza } H_0$$

Se dice que a nivel $\alpha = 0,05$ los datos muestran evidencia en contra de H_0 , o que los datos son significativos.

Test t para la diferencia de medias en poblacionales normales

Sea X_1, \dots, X_{n_1} una m.a.s. de $N(\mu_1, \sigma)$, e Y_1, \dots, Y_{n_2} una m.a.s. de $N(\mu_2, \sigma)$, ambas muestras independientes. Supondremos que μ_1 , μ_2 y σ son parámetros desconocidos. Se desea contrastar con nivel de significación α las siguientes hipótesis:

$$H_0 : \mu_1 - \mu_2 = K \quad \text{frente a} \quad H_1 : \mu_1 - \mu_2 \neq K$$

En general, el valor de K con mayor interés es $K = 0$, es decir, contrastar medias iguales frente a medias diferentes. La región crítica que proporciona el test de razón de verosimilitudes es:

$$C = \left\{ X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2} \left| \left| \frac{\bar{X} - \bar{Y} - K}{\sqrt{n_1 S_X^2 + n_2 S_Y^2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \sqrt{n_1 + n_2 - 2} > t_{n_1 + n_2 - 2, \frac{\alpha}{2}} \right\}$$

Ejemplo 10.22:

Si en el ejemplo anterior asumimos que las varianzas son desconocidas pero iguales, y que al observar la primera muestra se obtuvo $125 \text{ cm}^2 = S_X^2$, y en la segunda $169 \text{ cm}^2 = S_Y^2$, entonces, con estos datos, contrastar a nivel $\alpha = 0,05$ la igualdad de las medias poblacionales.

$$\frac{172 - 175 \sqrt{100 + 225 - 2}}{\sqrt{100 \times 125 + 225 \times 169} \sqrt{\frac{1}{100} + \frac{1}{225}}} = -1,9958$$

Como $t_{323, 0,05} = 1,96$, entonces los datos muestran evidencia en contra de H_0 .

Aunque en el ejemplo anterior se ha asumido la igualdad de varianzas, en las aplicaciones esa hipótesis debe ser contrastada.

Test F para la igualdad de varianzas en poblaciones normales

Sea X_1, \dots, X_{n_1} una m.a.s. de $N(\mu_1, \sigma_1)$, e Y_1, \dots, Y_{n_2} una m.a.s. de $N(\mu_2, \sigma_2)$, ambas muestras independientes. Supondremos que σ_1 y σ_2 son parámetros desconocidos. Se desea contrastar con nivel de significación α las siguientes hipótesis:

$$H_0 : \frac{\sigma_1}{\sigma_2} = 1 \quad \text{frente a} \quad H_1 : \frac{\sigma_1}{\sigma_2} \neq 1$$

La región crítica del test de razón de verosimilitudes de nivel de significación α es:

$$C = \left\{ X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2} \mid \frac{\hat{S}_X^2}{\hat{S}_Y^2} \geq F_{n_1-1, n_2-1, \frac{\alpha}{2}} \quad \text{ó} \quad \frac{\hat{S}_X^2}{\hat{S}_Y^2} \leq F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}} \right\}$$

Ejemplo 10.23

Siguiendo con el ejemplo anterior, contrastemos si realmente se puede admitir la igualdad de varianzas, para $n_1 = 100$, $n_2 = 225$, $S_X^2 = 125$ y $S_Y^2 = 169$.

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 \iff \sigma_1^2 = \sigma_2^2 \quad \text{frente a} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

$$\frac{\hat{S}_X^2}{\hat{S}_Y^2} = \frac{n_1(n_2-1)S_X^2}{n_2(n_1-1)S_Y^2} = \frac{100 \times 224 \times 125}{225 \times 99 \times 169} = 0,74$$

Si $\alpha = 0$, entonces $0,975 = p(F_{99,224} > a) = p\left(\frac{1}{F_{99,224}} < \frac{1}{a}\right) = p(F_{224,99} < \frac{1}{a}) \implies 0,025 = p(F_{224,99} > \frac{1}{a}) \implies \frac{1}{a} = 1,25 \implies a = \frac{1}{1,25} = 0,8$. Puesto que $0,74 < 0,8$, los datos muestran evidencia contra H_0 . Por tanto, la decisión tomada en el ejemplo anterior no es fiable, dado que se basa en la hipótesis de que las varianzas poblacionales eran iguales, y el test de igualdad de varianzas acaba de mostrar que dicha hipótesis no es aceptable.

No existe un test universalmente admitido para contrastar la diferencia de medias con varianzas desconocidas y distintas. Los programas informáticos suelen adoptar la hipótesis de Behrens-Fisher.

Test t para la diferencia de medias en poblaciones relacionadas y normales

Este test se utiliza cuando por la naturaleza del problema no tiene sentido utilizar muestras independientes, dado que en ese caso las diferencias observadas pudieran ser debidas a los individuos y no a las medias.

Ejemplo 10.24

Si se desea contrastar si dos crecepelos son, en media, igualmente efectivos, deben ser ambos aplicados a los mismos individuos. En otro caso, las diferencias observadas podrían ser debidas a los individuos y no a los crecepelos.

Entonces, sean $(X_1, Y_1), \dots, (X_n, Y_n)$ las observaciones correspondientes a n individuos. Asumimos que $X_i \sim N(\mu_1, \sigma_1)$, y $Y_i \sim N(\mu_2, \sigma_2)$.

Se desea contrastar $H_0 : \mu_1 = \mu_2$ frente a $H_1 : \mu_1 \neq \mu_2$.

Para la ejecución consideramos las nuevas observaciones, $D_1 = X_1 - Y_1, \dots, D_n = X_n - Y_n$. A estas nuevas variables se les aplica el test t para la media con varianzas desconocidas. La región crítica $C = \left\{ D_1, \dots, D_n \mid \left| \frac{\bar{D}-0}{S_D} \sqrt{n-1} \right| > t_{n-1, \frac{\alpha}{2}} \right\}$.

10.7 Interpretación de los resultados proporcionados por los paquetes estadísticos. Concepto de p -valor

Los resultados de los paquetes estadísticos no hacen referencia al concepto de región crítica ni de nivel de significación; la información se basa en el concepto de p -valor (o significación) que

se define como la probabilidad de observar un valor del estadístico de contraste más extremo que el que realmente se ha observado.

Ejemplo 10.25

En el ejemplo del test Z para la media de una normal con varianza conocida y datos acerca de latas de sardinas, si $n = 100$ latas, $H_0 : \mu = 1000$ frente a $\mu \neq 1000$, $\sigma_0 = 5$, $C = \left\{ \left| \frac{\bar{X} - 1000}{5} \right| > z_{\frac{\alpha}{2}} \right\}$ a nivel $\alpha = 0,05$, $z_{\frac{\alpha}{2}} = 1,96$ y rechazaremos si $\bar{X} > 100,98$ ó $\bar{X} < 999,02$.

Supongamos que $\bar{X} = 1003,21$ La información proporcionada por el ordenador sería: $\frac{\bar{X} - 1000}{\sigma_0} \sqrt{n} = \frac{1003,21 - 1000}{5} \sqrt{100} = 6,4$ y $p\text{-valor} = p(|Z| > 6,42) = 0$.

Rechazaremos H_0 para todo $\alpha > p\text{-valor}$. Si $\alpha > p\text{-valor}$, entonces el valor observado, en este caso, 6.42, cae dentro de la región crítica, pues genera unas colas de probabilidad más pequeñas que el extremo $z_{\frac{\alpha}{2}}$ correspondiente a nivel α .

Bibliografía

1. Neyman, J. y Pearson, E. (1928). *On the use and interpretation of certain test criteria for purposes of statistical inference: part I*. Biometrika, 20(1/2):175-242.
2. Neyman, J. (1977). *Frequentist probability and frequentist statistics*. Synthese, 36:97-131.
3. Fisher, R. (1925). *Theory of Statistical Estimation*. Mathematical Proceedings of the Cambridge Philosophical Society, 22(5):700-725.
4. Fisher, R. (1935). *The Logic of Inductive Inference*. Journal of the Royal Statistical Society, 39-82.
5. Butler, C. (1985). *Statistics in Linguistics*. Basil Blackwell.

Tema 11

Estimación por intervalos. Introducción. Métodos para encontrar estimadores de intervalos: método de la cantidad pivotal, método general o de Neyman, intervalos de confianza de longitud mínima

11.1 Intervalos de confianza. Introducción

Un intervalo de confianza es una técnica de estimación paramétrica utilizada en inferencia estadística que permite delimitar un par de valores, dentro de los cuales se encontrará el parámetro desconocido con una confianza alta elegida por el investigador.

Definición 1. Sea X_1, X_2, \dots, X_n una muestra aleatoria simple (m.a.s.) de $X \sim f(x, \theta)$, siendo θ un parámetro desconocido. Llamaremos intervalo de confianza de nivel $1 - \alpha$ para θ a (U, V) tal que

$$p(U < \theta < V) = 1 - \alpha \quad (11.1)$$

Observaciones:

- U y V son estadísticos (dependen de la muestra y sólo de la muestra). En la práctica, son estimadores por defecto y por exceso de θ , de manera que U infraestima θ y V lo sobrevalora.
- El nivel de confianza lo fija el investigador. Por defecto es 0,95. Otros valores muy utilizados son 0,9 y 0,99. Al incrementar el nivel de confianza, el intervalo se hace más largo, perdiéndose así precisión.
- La interpretación del concepto de intervalo es frecuentista en el siguiente sentido. Por ejemplo, si se dispone de 100 realizaciones muestrales, habría 100 intervalos numéricos, $(U, V), (U^1, V^1), \dots, (U^{99}, V^{99})$.

Si $1 - \alpha = 0,95$, al menos 95 de esos 100 intervalos contendrán el verdadero valor de θ .

Se dispone de dos métodos para obtener intervalos de confianza:

1. Método de la función pivote (o de la cantidad pivotal).
2. Método de Neyman.

11.2 Método de la función pivote o cantidad pivotal

Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim f(x, \theta)$, siendo θ un parámetro desconocido. Se desea obtener un intervalo de confianza para θ de nivel $1 - \alpha$. Supongamos que existe una función de

la muestra y de θ cuya distribución no depende de θ . A esta función se le denomina **función pivote**, lo que da nombre al método. Sea $h(X_1, X_2, \dots, X_n, \theta)$ dicha función.

Existirán constantes K_1 y K_2 (que no dependerán de θ porque la distribución de $h(X_1, X_2, \dots, X_n, \theta)$ no depende de θ , tales que:

$$p(K_1 \leq h(X_1, X_2, \dots, X_n, \theta) \leq K_2) = 1 - \alpha$$

Despejando θ en dichas inecuaciones, se obtienen los límites de confianza para θ .

Este método es especialmente apropiado para obtener intervalos de confianza para parámetros desconocidos en poblaciones normales. Analicemos estas situaciones.

11.2.1 Intervalo de confianza para la media de una distribución normal con varianza conocida

Sea X_1, X_2, \dots, X_n una m.a.s. de $N(\mu, \sigma_0)$, siendo la media μ desconocida y la desviación típica σ_0 conocida. Se desea obtener un intervalo de confianza para μ de nivel $1 - \alpha$. Sabemos por el Teorema de Fisher que $\bar{X} \sim N\left(\mu, \frac{\sigma_0}{\sqrt{n}}\right)$. Por tanto, $\frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}}$ es una función pivote para μ , ya que depende de la muestra y de μ , pero su distribución no depende de μ .

Por tanto, existe $z_{\frac{\alpha}{2}}$ tal que

$$p\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha.$$

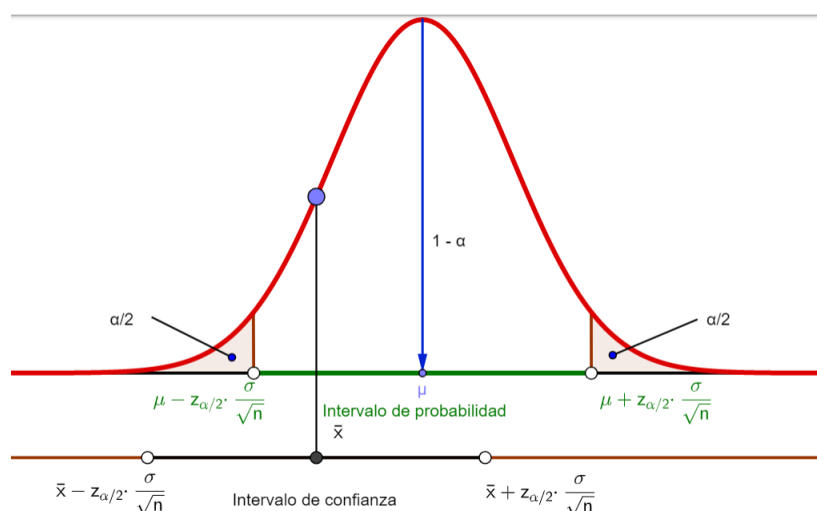
Entonces

$$-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}} \iff \bar{X} - \mu \geq -z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \iff \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}$$

y

$$\frac{\bar{X} - \mu}{\sigma_0/\sqrt{n}} \leq z_{\frac{\alpha}{2}} \iff \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \leq \mu.$$

Por tanto, el intervalo de confianza, de nivel $1 - \alpha$, es $\left(\bar{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}\right)$.

Figura 11.1: Intervalo de confianza $1 - \alpha$.

Como se puede observar, se han escogido los valores $-z_{\frac{\alpha}{2}}$ y $z_{\frac{\alpha}{2}}$ como valores críticos (simétricos) en la distribución $N(0, 1)$. Esta elección podría parecer inocente pero no lo es. Cualquier otra elección de valores críticos no simétricos, pero que dejen entre ambas colas de la $N(0, 1)$ una masa de α , garantizaría igualmente que el intervalo tenga nivel de confianza $1 - \alpha$. Cálculos sumamente engorrosos que tratamos de evitar aquí, demuestran que la elección de valores simétricos $z_{\frac{\alpha}{2}}$ y $-z_{\frac{\alpha}{2}}$ garantiza que la longitud del intervalo obtenido es mínima. Esta argumentación se extiende a todos los intervalos que se obtendrán a continuación a partir del método de la función pivote para la media de una normal o para la diferencia de medias en poblaciones normales. Para lectores que deseen conocer la metodología matemática según la cual se prueba que la longitud del intervalo es mínima, les remitimos a los ejemplos de intervalos de confianza obtenidos posteriormente a partir del método de Neyman, donde se ilustran estos cálculos.

Ejemplo 11.1:

Un proveedor de latas de sardinas afirma en el envase que el peso medio de las latas es de 100 gramos. Elegidas 36 latas al azar, el peso medio de la muestra ha resultado ser 96 gramos. Si se supone normalidad en el peso y varianza teórica (poblacional) igual a 9 gramos, obtener un intervalo del 95 % y otro del 99 % para la media poblacional, μ , con σ_0 conocido.

- Si $1 - \alpha = 0,95$, entonces $\alpha = 0,05$, con lo que $\frac{\alpha}{2} = 0,025$.

El intervalo de confianza numérico será $\left(96 \pm 1,96 \frac{3}{\sqrt{36}}\right) = (95,01, 96,98)$

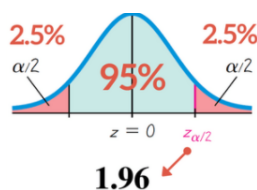


Figura 11.2: Intervalo al 95 % de confianza.

- Si $1 - \alpha = 0,99$, entonces $\alpha = 0,01$, con lo que $\frac{\alpha}{2} = 0,005$.

El intervalo de confianza numérico será $\left(96 \pm 2,57 \frac{3}{\sqrt{36}}\right) = (94,715, 97,528)$.

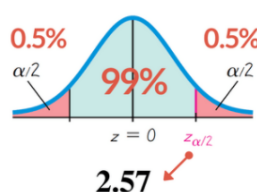


Figura 11.3: Intervalo al 99 % de confianza.

Se puede observar que a mayor confianza, menor es la precisión: el intervalo de confianza al 99 % es más ancho que el intervalo al 95 %.

11.2.2 Intervalo de confianza para la media de una normal con varianza desconocida

Sea X_1, X_2, \dots, X_n una m.a.s. de $N(\mu, \sigma)$, con μ y σ desconocidas. Se desea obtener un intervalo de nivel $1 - \alpha$ para μ .

Por el teorema de Fisher se tiene que:

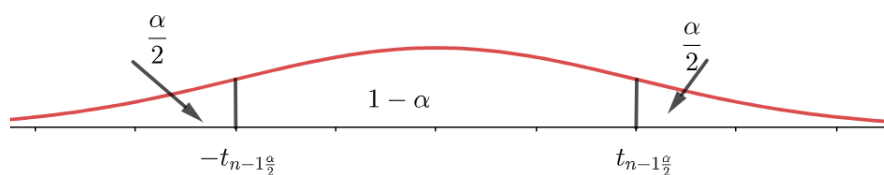
- $\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \rightsquigarrow N(0, 1)$.
- $\sqrt{\frac{nS^2}{\sigma^2(n-1)}} \rightsquigarrow \sqrt{\frac{\chi_{n-1}^2}{n-2}}$.

Y entonces, como en la expresión siguiente el numerador y el denominador son independientes,

$$\frac{\frac{\bar{X} - \mu}{\sigma} \sqrt{n}}{\sqrt{\frac{nS^2}{\sigma^2(n-1)}}} = \frac{\bar{X} - \mu}{S} \sqrt{n-1} \rightsquigarrow t_{n-1} \quad (11.2)$$

Esta es la función pivote. Existe $t_{n-1, \frac{\alpha}{2}}$ tal que

$$P\left(-t_{n-1, \frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{S} \sqrt{n-1} \leq t_{n-1, \frac{\alpha}{2}}\right) = 1 - \alpha$$



Despejando, tendríamos

$$-t_{n-1, \frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{S} \sqrt{n-1} \iff \mu \leq \bar{X} + \frac{t_{n-1, \frac{\alpha}{2}} S}{\sqrt{n-1}}$$

$$\frac{\bar{X} - \mu}{S} \sqrt{n-1} \leq t_{n-1, \frac{\alpha}{2}} \iff \mu \geq \bar{X} - \frac{t_{n-1, \frac{\alpha}{2}} S}{\sqrt{n-1}}$$

Por lo tanto, el intervalo resultante es:

$$\left(\bar{X} \pm \frac{t_{n-1, \frac{\alpha}{2}} S}{\sqrt{n-1}} \right) \quad (11.3)$$

Si se escribe el intervalo en términos de \hat{S} , como

$$S^2 = \frac{n-1}{n} \hat{S}^2 \implies S = \frac{\sqrt{n-1}}{\sqrt{n}} \hat{S}$$

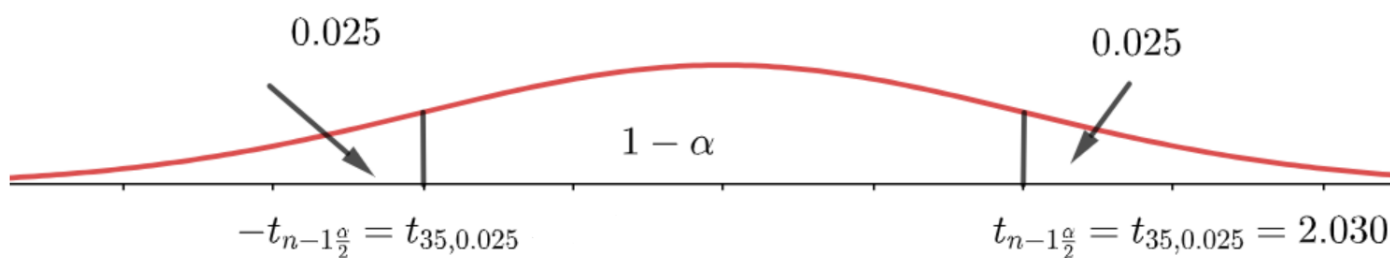
luego el intervalo es:

$$\left(\bar{X} \pm t_{n-1, \frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}} \right). \quad (11.4)$$

Ejemplo 11.2:

Si en el ejemplo anterior se elimina el conocimiento sobre la varianza, el intervalo al 95 % suponiendo que la cuasi-varianza del peso de las 36 latas es 16, se calcula como sigue.

Sabiendo que $\mu \in \left(\bar{X} \pm t_{n-1, \frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}} \right)$, y que $1 - \alpha = 0,95 \implies \alpha = 0,05 \implies \frac{\alpha}{2} = 0,025$, por lo que el intervalo al 95 % será $\left(96 \pm 2,03 \frac{4}{\sqrt{36}} \right)$.

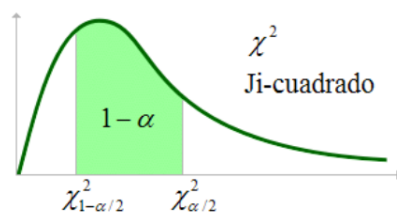


Si se quiere calcular el intervalo de confianza al 99 %, se tendría que $t_{35, 0.005} = 2,7281$, con lo que el intervalo será $\left(96 \pm 2,7281 \frac{4}{\sqrt{36}}\right)$.

11.2.3 Intervalo de confianza para la varianza de una normal conocida la media

Supongamos que X_1, X_2, \dots, X_n es una m.a.s. de $N(\mu_0, \sigma)$, μ_0 conocida y σ desconocida. Se desea calcular un intervalo de nivel de confianza $1 - \alpha$ para σ^2 .

$$\sum_{i=1}^n \left(\frac{X_i - \mu_0}{\sigma} \right)^2 \rightsquigarrow \chi_n^2 \iff \sum_{i=1}^n \frac{(X_i - \mu_0)^2}{\sigma^2} \rightsquigarrow \chi_n^2 \quad (11.5)$$



Así se obtiene una función pivote para σ^2 , luego existen $\chi_{n, \frac{\alpha}{2}}^2$ y $\chi_{n, 1-\frac{\alpha}{2}}^2$, tales que

$$P \left(\chi_{n, 1-\frac{\alpha}{2}}^2 \leq \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^2} \leq \chi_{n, \frac{\alpha}{2}}^2 \right) = 1 - \alpha.$$

Despejando,

$$\chi_{n, 1-\frac{\alpha}{2}}^2 \leq \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^2} \iff \sigma^2 \leq \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n, 1-\frac{\alpha}{2}}^2}$$

$$\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^2} \leq \chi_{n, \frac{\alpha}{2}}^2 \iff \sigma^2 \geq \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n, \frac{\alpha}{2}}^2}$$

luego el intervalo de nivel de confianza $1 - \alpha$ para σ^2 es

$$\left(\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n, \frac{\alpha}{2}}^2}, \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n, 1 - \frac{\alpha}{2}}^2} \right) \quad (11.6)$$

Lógicamente, para obtener un intervalo de nivel $1 - \alpha$ para σ , es suficiente con calcular la raíz cuadrada en los límites de confianza anteriores.

Ejemplo 11.3:

Supongamos que 1, 2, 3 es una realización muestral de una $N(\mu_0, \sigma)$, con $\mu_0 = 2$. Obtener un intervalo al 95 % para σ^2 .

$$\sum_{i=1}^n (X_i - \mu_0)^2 = (1 - 2)^2 + (2 - 2)^2 + (3 - 2)^2 = 2.$$

Por tanto, siendo $\chi_{3,0,975}^2 = 0,2157$ y $\chi_{3,0,025}^2 = 9,3484$ el intervalo de confianza al 95 % es $\left(\frac{2}{9,3484}, \frac{2}{0,2157} \right)$.

11.2.4 Intervalo de confianza para la varianza de una normal. Media desconocida

Supongamos que X_1, X_2, \dots, X_n es una m.a.s. de $N(\mu, \sigma)$, con μ y σ desconocidas. Se desea un intervalo de nivel $1 - \alpha$ para σ^2 .

Por el teorema de Fisher, $\frac{nS^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2$, que es entonces función pivote para σ^2 . Por tanto, existen $\chi_{n-1, \frac{\alpha}{2}}^2$ y $\chi_{n-1, 1 - \frac{\alpha}{2}}^2$ tales que

$$P \left(\chi_{n-1, 1 - \frac{\alpha}{2}}^2 \leq \frac{nS^2}{\sigma^2} \leq \chi_{n-1, \frac{\alpha}{2}}^2 \right) = 1 - \alpha.$$

Despejando,

$$\begin{aligned} \chi_{n-1, 1 - \frac{\alpha}{2}}^2 \leq \frac{nS^2}{\sigma^2} &\iff \sigma^2 \leq \frac{nS^2}{\chi_{n-1, 1 - \frac{\alpha}{2}}^2} \\ \frac{nS^2}{\sigma^2} \leq \chi_{n-1, \frac{\alpha}{2}}^2 &\iff \sigma^2 \geq \frac{nS^2}{\chi_{n-1, \frac{\alpha}{2}}^2}. \end{aligned}$$

Por tanto, el intervalo de nivel $1 - \alpha$ es

$$\left(\frac{nS^2}{\chi_{n-1, \frac{\alpha}{2}}^2}, \frac{nS^2}{\chi_{n-1, 1 - \frac{\alpha}{2}}^2} \right) \quad (11.7)$$

En realidad, la diferencia entre el caso anterior y éste es que si μ no es conocido, se estima con \bar{X} , con lo que se pierde un grado de libertad.

11.2.5 Intervalo de confianza para la diferencia de medias en poblaciones normales. Varianzas conocidas

Sea X_1, X_2, \dots, X_{n_1} una m.a.s. de $N(\mu_1, \sigma_1^2)$, con μ_1 desconocida y σ_1^2 conocida. A su vez, sea Y_1, Y_2, \dots, Y_{n_2} una m.a.s. de $N(\mu_2, \sigma_2^2)$, con μ_2 desconocida y σ_2^2 conocida. Además, se supone que ambas muestras son independientes. Se desea obtener un intervalo de confianza de nivel $1 - \alpha$ para $\mu_1 - \mu_2$. Sabemos que $\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$ y que $\bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$, y que ambas variables son independientes, puesto que las muestras correspondientes lo son. Entonces,

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) \iff \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1) \quad (11.8)$$

Por tanto, existe $z_{\frac{\alpha}{2}}$ tal que

$$p\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha \quad (11.9)$$

Despejando en las inecuaciones:

$$\begin{aligned} -z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} &\iff \mu_1 - \mu_2 \leq \bar{X} - \bar{Y} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \\ z_{\frac{\alpha}{2}} \geq \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} &\iff \mu_1 - \mu_2 \geq \bar{X} - \bar{Y} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \end{aligned}$$

luego el intervalo de nivel $1 - \alpha$ para $\mu_1 - \mu_2$ es

$$\left(\bar{X} - \bar{Y} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) \quad (11.10)$$

Ejemplo 11.4:

En el año 1980 la altura media de 1000 reclutas fue 1,7m, y en el año 1990 al altura media de 1021 reclutas fue de 1,73. Si se supone que las estaturas estaban normalmente distribuidas con varianzas respectivas de 15 y 17 cm^2 . Obtener un intervalo de confianza al 95 % para $\mu_1 - \mu_2$.

Puesto que $1 - \alpha = 0,95$, $\alpha = 0,05$ y $\frac{\alpha}{2} = 0,025$, el intervalo será

$$\left(170 - 173 \pm 1,96 \sqrt{\frac{15}{1000} + \frac{17}{1021}}\right) = (-3 \pm 0,35) = (-3,35, -2,65)$$

Puesto que $\mu_1 - \mu_2 = 0$ no pertenece al intervalo al 95 %, existe evidencia estadística de que la estatura media se ha modificado. Como además el intervalo tiene ambos extremos negativos, podemos concluir que $\mu_1 - \mu_2 < 0$, y por tanto, la estatura media (poblacional) de los reclutas se ha incrementado en esos 10 años.

11.2.6 Intervalo de confianza para la diferencia de medias en poblaciones normales con varianzas desconocidas pero iguales

Sea X_1, X_2, \dots, X_{n_1} una m.a.s. de $N(\mu_1, \sigma)$, e Y_1, Y_2, \dots, Y_{n_2} una m.a.s. de $N(\mu_2, \sigma)$, ambas muestras independientes, y μ_1, μ_2, σ desconocidos. Se desea obtener un intervalo de confianza para $\mu_1 - \mu_2$ de nivel $1 - \alpha$.

Sabemos que $\frac{n_1 S_X^2}{\sigma^2} \sim \chi_{n_1-1}^2$, y que $\frac{n_2 S_Y^2}{\sigma^2} \sim \chi_{n_2-1}^2$, ambas independientes. Entonces, $\frac{n_1 S_X^2}{\sigma^2} + \frac{n_2 S_Y^2}{\sigma^2} \sim \chi_{n_1+n_2-2}^2$.
 $\chi_{n_1-1}^2 + \chi_{n_2-1}^2 \equiv \Gamma\left(\frac{n_1-1}{2}, \frac{1}{2}\right) + \Gamma\left(\frac{n_2-1}{2}, \frac{1}{2}\right) \equiv \Gamma\left(\frac{n_1+n_2-2}{2}, \frac{1}{2}\right) \sim \chi_{n_1+n_2-2}^2$

La distribución χ^2 hereda la reproductividad de la Gamma, y si sumamos dos distribuciones χ^2 independientes obtendremos otra con grados de libertad la suma de los respectivos grados de ambas. Además, $\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}\right)$, y por tanto:

$$\left\{ \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \right\} \sim N(0, 1)$$

$$\left\{ \frac{\sqrt{\frac{n_1 S_X^2 + n_2 S_Y^2}{\sigma^2(n_1+n_2-2)}}}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \right\} \sim \sqrt{\frac{\chi_{n_1+n_2-2}^2}{n_1+n_2-2}}$$

ambos, numerador y denominador, independientes por el teorema de Fisher, es decir,

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1 S_X^2 + n_2 S_Y^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sqrt{n_1+n_2-2}$$

sigue una $t_{n_1+n_2-2}$. Es la función pivote. Por tanto, existe $t_{n_1+n_2-2, \frac{\alpha}{2}}$ tal que

$$P\left(-t_{n_1+n_2-2, \frac{\alpha}{2}} \leq \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1 S_X^2 + n_2 S_Y^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq t_{n_1+n_2-2, \frac{\alpha}{2}}\right) = 1 - \alpha$$

Despejando se obtendrán los límites y el intervalo de confianza de nivel $1 - \alpha$,

$$\left(\bar{X} - \bar{Y} \pm t_{n_1+n_2-2, \frac{\alpha}{2}} \frac{\sqrt{\frac{n_1 S_X^2 + n_2 S_Y^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}{\sqrt{n_1+n_2-2}} \right). \quad (11.11)$$

Escrito en términos de las cuasi-varianzas, será igual a:

$$\mu_1 - \mu_2 \in \left(\bar{X} - \bar{Y} \pm t_{n_1+n_2-2, \frac{\alpha}{2}} \frac{\sqrt{(n_1-1)\widehat{S}_X^2 + (n_2-1)\widehat{S}_Y^2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}{\sqrt{n_1+n_2-2}} \right) \quad (11.12)$$

Ejemplo 11.5:

Una muestra de estatura de 10 niños arrojó una media de 97cm y una cuasi-desviación típica de 11cm. En una muestra de estatura de 20 niñas se obtuvo una media de 112cm y cuasi-desviación típica de 17cm. Calcular un intervalo de confianza para la diferencia de estaturas medias niños y niñas.

Se tiene que $1 - \alpha = 0,95$, por lo que $\frac{\alpha}{2} = 0,05$, de donde $t_{28,0,025} = 2,043$. Por tanto, el intervalo de confianza buscado es

$$\left(97 - 112 \pm 2,043 \frac{\sqrt{1089 + 5491} \sqrt{\frac{1}{10} + \frac{1}{20}}}{\sqrt{28}} \right) = (-15 \pm 12,13) = (-27,13, -2,87).$$

11.2.7 Intervalo de confianza para el cociente de varianzas en poblaciones normales

Sea X_1, X_2, \dots, X_{n_1} una m.a.s. de $N(\mu_1, \sigma_1)$ y Y_1, Y_2, \dots, Y_{n_2} una m.a.s. de $N(\mu_2, \sigma_2)$, ambas muestras independientes, $\mu_1, \mu_2, \sigma_1, \sigma_2$ desconocidas. Se desea obtener un intervalo de confianza de nivel $1 - \alpha$ para $\frac{\sigma_1^2}{\sigma_2^2}$.

Calculamos un intervalo para el cociente y no para la diferencia, porque existe una función pivote para cociente y no para diferencia. Sin embargo, si nuestro interés es analizar la igualdad de varianzas, ello equivale a que el cociente sea 1, y por tanto, con un intervalo para el cociente se podría abordar el problema. Sabemos que $\frac{n_1 S_X^2}{\sigma_1^2} \sim \chi_{n_1-1}^2$, y que $\frac{n_2 S_Y^2}{\sigma_2^2} \sim \chi_{n_2-1}^2$ por el teorema de Fisher, siendo ambas variables independientes. Entonces,

$$\frac{\frac{n_1 S_X^2}{\sigma_1^2 (n_1-1)}}{\frac{n_2 S_Y^2}{\sigma_2^2 (n_2-1)}} \sim \frac{\chi_{n_1-1}^2}{\chi_{n_2-1}^2}$$

es decir,

$$\frac{n_1(n_2-1)S_X^2\sigma_2^2}{n_2(n_1-1)S_Y^2\sigma_1^2} \sim F_{n_1-1, n_2-1}.$$

Existen por tanto $F_{n_1-1, n_2-1, \frac{\alpha}{2}}$ y $F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}}$ tales que

$$p \left(F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}} \leq \frac{n_1(n_2-1)S_X^2\sigma_2^2}{n_2(n_1-1)S_Y^2\sigma_1^2} \leq F_{n_1-1, n_2-1, \frac{\alpha}{2}} \right) = 1 - \alpha$$

Despejando, el intervalo será

$$\frac{\sigma_1^2}{\sigma_2^2} \in \left(\frac{n_1(n_2-1)S_X^2}{n_2(n_1-1)S_Y^2 F_{n_1-1, n_2-1, \frac{\alpha}{2}}}, \frac{n_1(n_2-1)S_X^2}{n_2(n_1-1)S_Y^2 F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}}} \right).$$

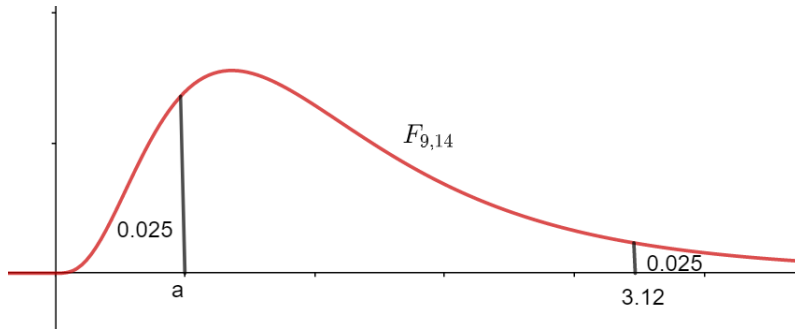
Si se desea escribir en términos de las cuasi-varianzas; como $\frac{n_1 S_X^2}{n_1-1} = \hat{S}_X^2$ y $\frac{n_2 S_Y^2}{n_2-1} = \hat{S}_Y^2$, entonces:

$$\frac{\sigma_1^2}{\sigma_2^2} \in \left(\frac{\hat{S}_X^2}{\hat{S}_Y^2 F_{n_1-1, n_2-1, \frac{\alpha}{2}}}, \frac{\hat{S}_X^2}{\hat{S}_Y^2 F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}}} \right).$$

Ejemplo 11.6:

La cuasi-varianza del peso de una muestra de 10 naranjas es de 81 gramos², y la cuasi-varianza en una muestra de 15 limones es de 75 gramos². Calcular un intervalo de confianza de nivel 0.95 para el cociente de varianzas poblacionales del peso de naranjas y limones. Se supone normalidad.

$$\left(\frac{81}{75 \times 3,12}, \frac{81}{75 \times a} \right)$$



$$0,025 = p(F_{9,14} < a) = p\left(F_{14, 9} > \frac{1}{a}\right) \Rightarrow \frac{1}{a} = 3,798 \Rightarrow a = 0,26$$

El intervalo es (0,346, 4,154). Como 1 pertenece al intervalo al 95 % de confianza, podemos asumir varianzas iguales.

11.3 Método general o de Neyman

Supongamos que se dispone de una m.a.s. X_1, X_2, \dots, X_n de una población $X \sim f(x, \theta)$, siendo θ un parámetro desconocido. Se desea obtener un intervalo de confianza para θ de nivel $1 - \alpha$. Si el método de la función pivote no fuera utilizable porque no existiera una función pivote, o porque fuera desconocida, se utiliza el método general, también conocido como método de Neyman, que no es más que una generalización del método de la función pivote.

Cuando se aplica el método general, partimos del estimador de máxima verosimilitud (EMV) de θ , $\hat{\theta}$, y supongamos que su distribución depende de θ . Entonces, existirán dos cantidades, $g_1(\theta)$ y $g_2(\theta)$ que dependen de θ tales que $p(\hat{\theta} \leq g_1(\theta)) = \alpha_1$, y $p(\hat{\theta} > g_2(\theta)) = \alpha_2$, con $\alpha_1, \alpha_2 \geq 0$ y $\alpha_1 + \alpha_2 = \alpha$.

Despejando en estas expresiones, podemos obtener los límites de confianza para θ .

En realidad, con este procedimiento obtenemos una infinidad de intervalos para θ , uno por cada elección de α_1 y α_2 . Se requiere entonces un criterio adicional para elegir entre estos infinitos intervalos uno concreto. Tal criterio es determinar, entre todos los posibles intervalos, el más corto (lógicamente, a igualdad de nivel de confianza, preferimos el intervalo más corto posible, porque es más preciso).

Ejemplo 11.7:

Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim U(0, \theta)$. Obtener el intervalo de confianza más corto para θ de nivel $1 - \alpha$.

Como no se conoce ninguna función pivote, se aplica el método de Neyman. Aquí, $\hat{\theta} = \max\{X_1, X_2, \dots, X_n\}$, con

$$f_{\hat{\theta}}(t) = \begin{cases} \frac{nt^{n-1}}{\theta^n}, & \text{si } t \in (0, \theta) \\ 0, & \text{en otro caso} \end{cases}$$

Fijamos $\alpha_1, \alpha_2 \in [0, 1]$, tales que $\alpha_1 + \alpha_2 = \alpha$.

Imponiendo la condición $\alpha_1 = p(\hat{\theta} \leq g_1(\theta)) = \int_0^{g_1(\theta)} \frac{nt^{n-1}}{\theta^n} dt = \frac{t^n}{\theta^n} \Big|_0^{g_1(\theta)} = \frac{g_1(\theta)^n}{\theta^n} \implies g_1(\theta) = \theta \sqrt[n]{\alpha_1}$.

Imponiendo la otra condición, $\alpha_2 = p(\hat{\theta} \geq g_2(\theta)) = \int_{g_2(\theta)}^{\theta} \frac{nt^{n-1}}{\theta^n} dt = \frac{t^n}{\theta^n} \Big|_{g_2(\theta)}^{\theta} = 1 - \frac{g_2(\theta)^n}{\theta^n} \implies \frac{g_2(\theta)^n}{\theta^n} = 1 - \alpha_2 \implies g_2(\theta) = \theta \sqrt[n]{1 - \alpha_2}$.

Por tanto,

$$p(g_1(\theta) \leq \hat{\theta} \leq g_2(\theta)) = 1 - \alpha \iff p(\theta \sqrt[n]{\alpha_1} \leq \hat{\theta} \leq \theta \sqrt[n]{1 - \alpha_2}) = 1 - \alpha.$$

Despejando,

$$\theta \sqrt[n]{\alpha_1} \leq \hat{\theta} \implies \theta \leq \frac{\hat{\theta}}{\sqrt[n]{\alpha_1}} \text{ y } \hat{\theta} \geq \theta \sqrt{1 - \alpha_2} \iff \theta \geq \frac{\hat{\theta}}{\sqrt{1 - \alpha_2}}$$

luego un intervalo de confianza para θ de nivel $1 - \alpha$ será:

$$\left(\frac{\hat{\theta}}{\sqrt{1 - \alpha_2}}, \frac{\hat{\theta}}{\sqrt[n]{\alpha_1}} \right) \quad (11.13)$$

En realidad, se han obtenido infinitos intervalos, uno para cada elección de α_1 y α_2 . Buscamos el más corto. La longitud es

$$L(\alpha_1) = \hat{\theta} \left(\frac{1}{\sqrt[n]{\alpha_1}} - \frac{1}{\sqrt{1 - \alpha - \alpha_1}} \right)$$

$$L'(\alpha_1) = \hat{\theta} \left(-\frac{1}{n} \alpha_1^{\frac{-1}{n}-1} + \frac{1}{n} (1 - \alpha - \alpha_1)^{\frac{-1}{n}-1} \right) = 0 \implies$$

$$\implies \frac{1}{n} (1 - \alpha - \alpha_1)^{\frac{-1}{n}-1} = \frac{1}{n} (\alpha_1)^{\frac{-1}{n}-1} \iff 1 - \alpha = 0 \text{ lo que es imposible}$$

Es decir, no hay raíces y $L'(\alpha_1) \neq 0$ siempre. Así que, $L(\alpha_1)$ es siempre creciente o siempre decreciente, $L(\alpha_1)$ está definida para $\alpha_1 \in [0, \alpha]$.

Como $L(0) = \infty$, L decrece y la longitud mínima se obtiene cuando $\alpha_1 = \alpha$ y $\alpha_2 = 0$, y el intervalo de confianza resultante es:

$$\left(\hat{\theta} = \max\{X_1, \dots, X_n\}, \frac{\hat{\theta}}{\sqrt[n]{\alpha}} \right)$$

Si $x_1 = 0,3$, $x_2 = 0,5$, $x_3 = 0,1$, el intervalo para θ de nivel 0.95 y longitud menor, cuando $X \sim U(0, \theta)$ es $\left(0,5, \frac{0,5}{\sqrt[3]{0,05}}\right)$

Ejemplo 11.8:

Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim f(x, \theta)$, con

$$f(x, \theta) = \begin{cases} e^{-x} e^{\theta}, & x > \theta \\ 0, & \text{en otro caso.} \end{cases}$$

Entonces $\hat{\theta} = \min\{X_1, \dots, X_n\}$ y

$$f_{\hat{\theta}}(t) = \begin{cases} n e^{n\theta} e^{-nt} & \text{si } t > \theta \\ 0 & \text{en otro caso} \end{cases}$$

Fijando $\alpha_1, \alpha_2 \in [0, 1]$, tales que $\alpha_1 + \alpha_2 = \alpha$, e imponiendo la condición

$$\alpha_1 = p\left(\widehat{\theta} \leq g_1(\theta)\right) = \int_{\theta}^{g_1(\theta)} ne^{n\theta} e^{-nt} dt = e^{n\theta} \left[-e^{-nt} \right]_{\theta}^{g_1(\theta)} = e^{n\theta} [-e^{-ng_1(\theta)} + e^{-n\theta}] =$$

$$-e^{n(\theta - g_1(\theta))} + 1 \implies 1 - \alpha_1 = e^{n(\theta - g_1(\theta))} \implies \ln(1 - \alpha_1) = n(\theta - g_1(\theta)) \implies$$

$$g_1(\theta) = \theta - \frac{1}{n} \ln(1 - \alpha_1)$$

Por otra parte, imponiendo

$$\alpha_2 = p\left(\widehat{\theta} \geq g_2(\theta)\right) = \int_{g_2(\theta)}^{\infty} ne^{-nt} dr = e^{n\theta} [-e^{-nt}]_{g_2(\theta)}^{\infty} = e^{n\theta} e^{-ng_2(\theta)};$$

$$e^{n(\theta - g_2(\theta))} \implies \ln \alpha_2 = n(\theta - g_2(\theta)) \implies \frac{1}{n} \ln \alpha_2 = \theta - g_2(\theta) \implies$$

$$g_2(\theta) = \theta - \frac{1}{n} \ln \alpha_2.$$

$$\text{Así, } 1 - \alpha = p\left(g_1(\theta) \leq \widehat{\theta} \leq g_2(\theta)\right) = p\left(\theta - \frac{1}{n} \ln(1 - \alpha_1) \leq \widehat{\theta} \leq \theta - \frac{1}{n} \ln \alpha_2\right).$$

Despejando,

$$\theta - \frac{1}{n} \ln(1 - \alpha_2) \leq \widehat{\theta} \implies \theta \leq \widehat{\theta} + \frac{1}{n} \ln(1 - \alpha_1)$$

$$\widehat{\theta} \geq \theta - \frac{1}{n} \ln \alpha_2 \implies \theta \geq \widehat{\theta} + \frac{1}{n} \ln \alpha_2$$

Un intervalo de nivel $1 - \alpha$ es $\left(\widehat{\theta} + \frac{1}{n} \ln \alpha_2, \widehat{\theta} + \frac{1}{n} \ln(1 - \alpha_1)\right)$.

En realidad, para cada elección de α_1 y α_2 hay un intervalo distinto. Elegiremos los valores de α_1 y α_2 que nos garanticen que el intervalo es el más corto posible, para que así sea el más preciso.

$$L(\alpha_1) = \frac{1}{n} \ln(1 - \alpha_1) - \frac{1}{n} \ln \alpha_2 = \frac{1}{n} \ln(1 - \alpha_1) - \frac{1}{n} \ln(\alpha - \alpha_1)$$

$$L'(\alpha_1) = \frac{1}{n} \frac{(-1)}{1 - \alpha_1} - \frac{1}{n} \frac{(-1)}{\alpha - \alpha_1} = 0 \implies \frac{1}{1 - \alpha_1} - \frac{1}{\alpha - \alpha_1} = 0 \implies \alpha - 1 = 0,$$

lo que no es posible.

Por tanto, $L'(\alpha_1)$ no se anula, y $L(\alpha_1)$ es creciente o decreciente siempre en $[0, 1]$. Como $L(\alpha) = +\infty$, entonces L es creciente de 0 a α y la longitud mínima se obtiene para $\alpha_1 = 0$ y $\alpha_2 = \alpha$, o sea,

$$\left(\hat{\theta} + \frac{1}{n} \ln \alpha, \hat{\theta} \right). \quad (11.14)$$

Ejemplo 11.9:

Si se tiene

$$f(x, \theta) = \begin{cases} e^{-x} e^{\theta}, & x > \theta \\ 0, & \text{en otro caso} \end{cases}$$

y $\hat{\theta} = 2,3$ y tres observaciones ($n = 3$), el intervalo de nivel 0,95 más corto para θ será

$$\left(2,3 + \frac{1}{3} \ln 0,05, 2,3 \right).$$

11.4 Intervalos para parámetros de distribuciones discretas

Las funciones de distribución de variables discretas presentan saltos en todos los valores que toman las variables. Por tanto, para la mayoría de los valores $1 - \alpha$ no se puede obtener un intervalo de dicho nivel. En tales casos, se utilizan tamaños muestrales grandes para, haciendo uso del teorema central del límite, obtener intervalos de confianza aproximados para los parámetros. De nuevo, el punto de partida será el EMV $\hat{\theta}$ que, en general es consistente y asintóticamente normal (CAN).

11.4.1 Intervalo de confianza aproximado para p en una población $B(1, p)$ si n es grande ($n \geq 30$).

Sabemos que $\hat{p} = \bar{X}$ y que aproximadamente $\bar{X} \rightsquigarrow N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$ si $n \geq 30$, luego $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow$ aproximadamente $N(0, 1)$.

Existe $z_{\frac{\alpha}{2}}$ tal que $p\left(-z_{\frac{\alpha}{2}} \leq \frac{\hat{p}-p}{\sqrt{p(1-p)}} \sqrt{n} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$.

Despejando,

$$\begin{aligned} \frac{\hat{p}-p}{\sqrt{p(1-p)}} \sqrt{n} \leq z_{\frac{\alpha}{2}} &\implies p \geq \hat{p} - \frac{z_{\frac{\alpha}{2}} \sqrt{p(1-p)}}{\sqrt{n}} \\ \frac{\hat{p}-p}{\sqrt{p(1-p)}} \sqrt{n} \geq -z_{\frac{\alpha}{2}} &\implies p \leq \hat{p} + \frac{z_{\frac{\alpha}{2}} \sqrt{p(1-p)}}{\sqrt{n}} \end{aligned}$$

Como los límites del intervalo de confianza dependen de p (desconocido), no nos sirven, y estimamos p por \hat{p} para evitar la incertidumbre, de modo que el intervalo será:

$$\left(\hat{p} \pm z_{\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \right). \quad (11.15)$$

Ejemplo 11.10:

Se lanza una moneda 50 veces y se obtienen 20 caras. Obtener un intervalo aproximado al 95 % para la probabilidad de cara de la moneda.

$$z_{\frac{\alpha}{2}} = 1,96 \quad \hat{p} = \frac{20}{50} = 0,4$$

$$\left(0,4 - 1,96 \frac{\sqrt{0,24}}{\sqrt{50}}, 0,4 + 1,96 \frac{\sqrt{0,24}}{\sqrt{50}} \right)$$

11.4.2 Intervalo de confianza aproximado para λ en el modelo de Poisson

Sea X_1, X_2, \dots, X_n una m.a.s. de $X \sim P(\lambda)$. Se desea obtener un intervalo de confianza aproximado para λ de nivel $1 - \alpha$. Supondremos que $n \geq 30$. Entonces, por el teorema central del límite, \bar{X} sigue aproximadamente $N\left(\lambda, \frac{\sqrt{\lambda}}{\sqrt{n}}\right)$. Por tanto, existe $z_{\frac{\alpha}{2}}$ tal que $p\left(-z_{\frac{\alpha}{2}} \leq N(0, 1) \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$. En particular, $\frac{\bar{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} \sim N(0, 1)$, luego $p\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$.

Despejando en las inecuaciones,

$$-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} \iff \lambda \leq \bar{X} + z_{\frac{\alpha}{2}} \sqrt{\frac{\lambda}{n}}$$

$$z_{\frac{\alpha}{2}} \geq \frac{\bar{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} \iff \lambda \geq \bar{X} - z_{\frac{\alpha}{2}} \sqrt{\frac{\lambda}{n}}$$

Siendo n grande, estimamos λ en los límites de confianza, que dependen de λ , a través de \bar{X} (consistente) para obtener el intervalo aproximado de nivel $1 - \alpha$.

$$\left(\bar{X} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{X}}{n}} \right) \quad (11.16)$$

Ejemplo 11.11:

Se desea estimar el número medio de personas que entran en un centro comercial los lunes entre

las 2 y las 3 de la tarde. Examinamos los primeros 50 lunes de un año en esa franja horaria, y se obtiene que entraron un total de 550 personas. Obtener un intervalo de confianza aproximado para la media teórica de visitas, al 95 % de confianza.

Supongamos que las entradas en el centro comercial se distribuyen según Poisson. Como $1 - \alpha = 0,95$, $z_{\frac{\alpha}{2}} = 1,96$. Aplicando la fórmula,

$$\left(\frac{550}{50} \pm 1,96 \sqrt{\frac{550}{2500}} \right)$$

es un intervalo aproximado de nivel 95 % para λ .

11.4.3 Intervalo de confianza para la diferencia de proporciones

Sea X_1, X_2, \dots, X_{n_1} una m.a.s de $B(1, p_1)$, y sea Y_1, Y_2, \dots, Y_{n_2} una m.a.s de $B(1, p_2)$, ambas muestras independientes. Admitiremos que $n_1 \geq 30$, y $n_2 \geq 30$. Se desea obtener un intervalo de confianza para $p_1 - p_2$ de nivel $1 - \alpha$.

Por el teorema central del límite, se sabe que $\hat{p}_1 = \bar{X} \rightsquigarrow N\left(p_1, \sqrt{\frac{p_1(1-p_1)}{n_1}}\right)$ aproximadamente, y que $\hat{p}_2 = \bar{Y} \rightsquigarrow N\left(p_2, \sqrt{\frac{p_2(1-p_2)}{n_2}}\right)$, aproximadamente.

Como las dos muestras son independientes, también lo van a ser las distribuciones de \hat{p}_1 y \hat{p}_2 . Entonces, de forma aproximada,

$$\hat{p}_1 - \hat{p}_2 \rightsquigarrow N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

o tipificado

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \rightsquigarrow N(0, 1), \quad \text{aproximadamente.}$$

Entonces, existe $z_{\frac{\alpha}{2}}$ tal que $p\left(-z_{\frac{\alpha}{2}} \leq N(0, 1) \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$. Por tanto,

$$p\left(-z_{\frac{\alpha}{2}} \leq \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha.$$

Despejando,

$$-z_{\frac{\alpha}{2}} \leq \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \iff p_1 - p_2 \leq \hat{p}_1 - \hat{p}_2 + z_{\frac{\alpha}{2}} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \leq z_{\frac{\alpha}{2}} \iff p_1 - p_2 \geq \hat{p}_1 - \hat{p}_2 - z_{\frac{\alpha}{2}} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Como los límites de confianza dependen de p_1 y p_2 , estimamos p_1 y p_2 a través de \hat{p}_1 y \hat{p}_2 , que son consistentes (cuanto mayor sea el tamaño de n_1 y n_2 , menor error se comete). El intervalo aproximado de nivel $1 - \alpha$ para $p_1 - p_2$ es:

$$\left(\hat{p}_1 - \hat{p}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \right). \quad (11.17)$$

Ejemplo 11.12:

Se desea obtener un intervalo de confianza aproximado de nivel 90 % para la diferencia entre la proporción de aprobados en las asignaturas de A y B. De 44 alumnos consultados, aprobaron 33 el examen final de A, y de 32 consultados aprobaron 18 el examen final de B. Supondremos que existe independencia entre las notas de ambos exámenes. Un intervalo del 90 % para $p_1 - p_2$ siendo p_1 la proporción de aprobados en A y p_2 la proporción de aprobados en B es:

$$\left(\frac{33}{44} - \frac{18}{32} \pm 1,645 \sqrt{\frac{33 \times 11}{44 \times 44 \times 44} + \frac{18 \times 14}{32 \times 32 \times 32}} \right) = (-0,0157, 0,3907)$$

Al 95 % de confianza, el intervalo obtenido sería

$$\left(\frac{33}{44} - \frac{18}{32} \pm 1,96 \sqrt{\frac{33 \times 11}{44 \times 44 \times 44} + \frac{18 \times 14}{32 \times 32 \times 32}} \right) = (-0,04, 0,42)$$

Bibliografía

1. Neyman, J. y Pearson, E. (1928). *On the use and interpretation of certain test criteria for purposes of statistical inference: part I*. Biometrika, 20(1/2):175-242.
2. Neyman, J. (1977). *Frequentist probability and frequentist statistics*. Synthese, 36:97-131.
3. Fisher, R. (1925). *Theory of Statistical Estimation*. Mathematical Proceedings of the Cambridge Philosophical Society, 22(5):700-725.
4. Fisher, R. (1935). *The Logic of Inductive Inference*. Journal of the Royal Statistical Society, 39-82.
5. Butler, C. (1985). *Statistics in Linguistics*. Basil Blackwell.

Tema 12

Estadística Descriptiva I. Concepto de estadística. Las unidades estadísticas. Variables cualitativas y cuantitativas. Variables discretas y continuas. Distribuciones unidimensionales de frecuencias. Distribuciones acumuladas. Tablas estadísticas. Representación gráfica.

12.1 Concepto de estadística

La estadística es una colección de métodos que nos ayudan a describir, resumir, interpretar, y analizar datos. Sacar conclusiones de los datos es vital en la investigación, la administración, y negocios. Por ejemplo, los investigadores pueden estar interesados en comprender si una intervención médica ayuda a reducir la carga de una enfermedad, a entender la relación entre la personalidad de los humanos y sus tomas de decisiones, si un nuevo fertilizante aumentaría el rendimiento de los cultivos, cómo un sistema político afecta a la política comercial, o quiénes van a ser los partidos más votados en unas elecciones. Los gobiernos y las organizaciones pueden estar interesados en la esperanza de vida de una población, en los factores de riesgo de mortalidad infantil, o los factores geográficos que afectan a posibles diferencias en el uso de energía, además de en el estudio de patrones de migración o razones del desempleo. Por otra parte, el interés de un negocio puede estar en la identificación de personas que puedan estar interesadas en un determinado producto, optimización de precios, y evaluar de la satisfacción de los clientes (Heumann y Shalabh 2016).

Independientemente de cuál sea su utilidad final, es muy importante recopilar datos de manera que su análisis posterior sea factible. Los datos recopilados nos permiten la aplicación de una amplia gama de métodos estadísticos. En este tema, se presentan las distintas herramientas estadísticas necesarias para recopilar, administrar, evaluar y analizar datos.

12.2 Las unidades estadísticas

Los elementos o unidades estadísticas son las entidades sobre las que se obtienen datos. Por ejemplo, 1) si se evalúa el conocimiento de idiomas de los aspirantes a un puesto de trabajo en una empresa, cada aspirante es una unidad. 2) Si se estudia el fenómeno de abandono escolar de los distintos colegios de una región, cada colegio de dicha región es una unidad. 3) Si a un vendedor le interesa el volumen de ventas mensual de su comercio en el último año, cada mes del año en cuestión es una unidad.

Las unidades estadísticas se caracterizan mediante variables descriptivas que permiten la identificación adecuada de la unidad de interés. Estas características permiten la recogida de información acerca de las unidades y sus estructuras. Además, establecen una base de muestreo para los estudios estadísticos que permite efectuar comparaciones y vínculos entre los datos de diferentes fuentes. Esta posibilidad reduce notablemente los errores durante el proceso de recolección de datos.

Las unidades estadísticas son llamadas también unidades experimentales, unidades de muestreo o unidades de observación.

En general, el objetivo de los estudios estadísticos es la generalización de las unidades observadas a un conjunto más grande que incluye todas las unidades existentes, aunque algunas de ellas no se observen directamente.

En los conjuntos de datos más simples, cada unidad se corresponde con un valor de los datos. Sin embargo, cuando aumenta la complejidad del conjunto de datos en estudio, hace falta llevar a cabo varias mediciones para cada unidad, para lo cual hay que tener en cuenta que las mediciones tomadas de un mismo individuo están relacionadas en cierta manera, en el sentido de que son más parecidas entre sí que con respecto a las tomadas en distintos individuos. No tener en cuenta este hecho puede provocar problemas en análisis posteriores.

12.3 Variables

Una vez que se ha especificado la población de interés para una pregunta de investigación específica, el siguiente paso es especificar los factores que resultan de interés sobre las observaciones o datos obtenidos. Cada una de las características particulares de estas observaciones se puede recopilar y representar mediante una variable estadística, X . En general, cualquier tipo de información de interés puede ser representada por una variable. Por ejemplo, si los datos se refieren a un conjunto de personas, una variable X podría describir el estado civil, o el género, o la edad, por mencionar sólo algunos aspectos. En general, hay más de una característica de interés. Para poder representarlas, se consideran varias variables, tantas como características se quieran representar, de manera que cada característica de interés sería representada mediante una variable, $X_i, i = 1, 2, \dots, p$. Volviendo al ejemplo sobre el conjunto de personas, se podría considerar que X_1 representa el estado civil, X_2 el género y X_3 la edad de cada una de las personas involucradas. En este escenario, cada observación toma un valor concreto para cada una de las variables involucradas, en este caso, X_1, X_2, X_3 . Por ejemplo, X_2 tomará un valor, que en este caso será *masculino* o *femenino*, dependiendo de la observación considerada. Formalmente, una variable estadística se define de la siguiente manera.

Definición 1.

Sea Ω una población, y sea S un conjunto de valores. Una variable X definida sobre la población Ω se caracteriza como

$$X : \Omega \longrightarrow S$$

$$\omega \mapsto x$$

Esta definición establece que una variable X toma un valor x para cada observación $\omega \in \Omega$, donde el número de valores posibles está contenido en el conjunto S .

Ejemplo 12.1:

1. Si X se refiere al género, los posibles valores de x están contenidos en $S = \{ \text{masculino}, \text{femenino} \}$. Para cada observación ω , el valor de x será *masculino* o *femenino*. Esta información se resume en la variable X .
2. Sea X el país de fabricación de un automóvil. Si cada observación ω se refiere a un automóvil, los posibles valores que puede tomar la variable $X := \text{país}$ quedan determinados por el conjunto $S = \{ \text{Italia}, \text{Corea del Sur}, \text{Alemania}, \text{Francia}, \text{India}, \text{China}, \text{Japón}, \text{EE.UU.}, \dots \}$.
3. Si la variable X se refiere a la edad de una persona, podemos establecer que esta podrá tomar cualquier valor entre 1 y 125, de manera que $S = \{1, 2, 3, \dots, 125\}$. Cada observación w se refiere a una persona, a la que la variable X asigna el valor x correspondiente a su edad.

12.3.1 Variables cualitativas y cuantitativas

Dependiendo de su naturaleza, las variables se pueden clasificar atendiendo a distintos criterios. En esta sección se presentan las variables cualitativas y las variables cuantitativas.

Las variables **cualitativas** son las variables que toman valores x que no se pueden ordenar en de forma lógica o natural. Los siguientes ejemplos son casos muy representativos de variables cualitativas.

- El color de pelo de una persona
- El nombre de un partido político.
- El tipo de transporte utilizado para ir a trabajar.
- El tipo de vivienda en la que habita una persona.

No hay ninguna manera de ordenar los distintos colores de pelo, de la misma manera que no se puede establecer un orden relevante entre los nombres de distintos partidos políticos.

Por el contrario, las variables **cuantitativas** representan cantidades que se pueden medir. Los valores que toman las variables cuantitativas se pueden ordenar de forma lógica y natural. Algunos ejemplos de variables cuantitativas son:

- El tamaño de una mesa.
- El precio de una casa.
- Número de horas de estudio semanales.
- El peso o la altura de una persona.

Para fines prácticos, en ocasiones se les asignan números a variables cualitativas, que facilitan los análisis de datos posteriores. Por ejemplo, si se considera la variable *género*, cada observación puede tomar el valor *masculino* o *femenino*. Sin embargo, se puede decidir asignar el número 1 a las mujeres y el número 0 a los hombres, y usar estos números en lugar de las categorías originales. No obstante, hay que tener en cuenta que esta decisión pudiera ser arbitraria, e igualmente se podría haber elegido 1 para hombre y 0 para mujer, o 2 para *hombre* y 10 para

mujer. Así, ni aún habiendo asignado un número a cada categoría, se tendría un orden lógico y natural sobre cómo organizar los valores *hombre* y *mujer*, de manera que el género sigue siendo una variable cualitativa, incluso después de usar números para codificarla.

12.3.2 Variables discretas y continuas

En esta sección se presentan las variables discretas y las variables continuas.

Las variables **discretas** son variables que únicamente pueden tomar un número contable o numerable de valores. Todas las variables cualitativas son discretas, como por ejemplo

- El color de ojos.
- La región de un país.
- Las facultades de una Universidad.

También las variables cuantitativas pueden ser discretas, como por ejemplo

- La talla de zapatos.
- El número de cursos finalizados en una titulación.

Estas variables cuantitativas son discretas, ya que el conjunto de valores que pueden tomar es numerable.

Las variables **continuas** son aquellas variables que pueden tomar infinitos valores distintos, no numerables. Algunos ejemplos de variables continuas son:

- El tiempo empleado para llegar a trabajar.
- La distancia entre dos planetas.
- El peso o la altura de una persona.
- El volumen de una piscina.

En ocasiones se dice que las variables continuas son aquellas variables que *miden* en lugar de *contar*. Esta caracterización es muy informal, pero ayuda a entender la diferencia entre variables discretas y variables continuas. El punto crucial es que las variables continuas pueden, en teoría, tomar un número infinito no numerable de valores; por ejemplo, se puede registrar la altura de una persona como 172cm. Sin embargo, la altura real de la cinta métrica puede ser de 172,3cm, cantidad que se redondeó a 172 cm. Tal vez un instrumento de medición más preciso hubiera proporcionado el valor 172,342cm. Sin embargo, la altura real de esta persona es un número con indefinidamente muchos lugares decimales como por ejemplo 172,342975328...cm. Sin importar que en ocasiones se limite la información que proporciona, toda variable que pueda tomar una cantidad infinita de valores se define como una variable continua.

12.3.3 Escalas

Las consideraciones previas indican que las distintas variables contienen diferentes cantidades y tipos de información. Una clasificación útil de estas consideraciones viene dada por el concepto

de escala de una variable. Este concepto ayuda en la identificación de los métodos que resultan más adecuados para cada escenario particular.

- **Escala nominal.** Los valores de una variable nominal no se pueden ordenar. Algunos ejemplos son el género de una persona (*hombre-mujer*) o el estado de una solicitud (*pendiente - no pendiente*).
- **Escala ordinal.** Los valores de una variable ordinal se pueden ordenar. Sin embargo, las diferencias entre estos valores no se puede interpretar de manera significativa. Por ejemplo, los posibles valores del nivel educativo (*ninguno - educación primaria - educación secundaria - título universitario*) se pueden ordenar de manera significativa, pero las diferencias entre estos valores no se puede interpretar. Asimismo, la satisfacción con un producto (*insatisfecho - satisfecho - muy satisfecho*) es una variable ordinal porque los posibles valores que toma esta variable se pueden ordenar, pero las diferencias entre *insatisfecho – satisfecho* y *satisfecho – muy satisfecho* no se pueden comparar de forma numérica.
- **Escala continua.** Los valores de una variable continua se pueden ordenar. Además, las diferencias entre estos valores se pueden interpretar de manera significativa. Por ejemplo, la altura de una persona se refiere a una variable continua porque los valores se pueden medir (170 cm, 171 cm, 172 cm,...), y las diferencias entre estos valores se pueden comparar (la diferencia entre 170 y 171 cm es igual a la diferencia entre 171 y 172 cm). A veces, la escala continua es dividida en subescalas que simplifican su interpretación y utilización. Algunas de las subescalas continuas son las siguientes.
 - **De intervalo.** En esta subescala se clasifican aquellas variables para las que únicamente se pueden interpretar las diferencias entre los valores, pero no las proporciones. Un ejemplo de esta escala sería la temperatura (medida en $^{\circ}\text{C}$): la diferencia entre 2°C y 4°C es 6°C , pero la relación de $\frac{4}{2} = 2$ no significa que 4°C sea el doble de frío que 2°C .
 - **De proporción.** En esta subescala se clasifican aquellas variables para las que se pueden interpretar tanto las diferencias como las proporciones. Un ejemplo es la velocidad: 60km/h es a 40km/h 20km/h más. Además, 60km/h es tres veces más rápido que 20km/h porque la relación entre ellos es $\frac{60}{20} = 3$.
 - **Escala absoluta.** La escala absoluta es la misma que la escala de razón, con la diferencia de que los valores se miden en unidades *naturales*. Un ejemplo de variable de escala absoluta es aquella que mide el número de semestres estudiados. En este escenario no se necesita una unidad artificial como km/h o $^{\circ}\text{C}$ para medir los valores; estos serán simplemente 1, 2, 3, etc.

12.3.4 Datos agrupados

En algunas circunstancias, puede darse el caso de que los datos estén disponibles únicamente forma resumida: en vez de conocer el valor original de una variable, únicamente se puede conocer la categoría o grupo al que pertenece el valor en cuestión. Algunos escenarios en que se puede dar esta situación son:

- Encuestas relacionadas con ingresos anuales suelen ser organizadas por grupos del tipo $[0€ - 20.000€)$, $[20.000€ - 30.000€)$, ..., $> 100.000€$.
- En una campaña electoral en la que hay muchos partidos políticos, aquellos con un número bajo de votantes a menudo se suelen resumir en una nueva categoría, *Otros*.
- Una compañía de seguros, en vez de capturar el número de reclamaciones realizadas por un cliente, puede controlar simplemente si el cliente en cuestión ha reclamado o no, para lo cual se usaría una variable que toman los valores *Sí* - *No*.

Si los datos están disponibles en forma agrupada, la variable en cuestión que resume este tipo de información suele llamarse *variable agrupada*. En ocasiones, estas variables también se conocen como *variables categóricas*. Sin embargo, esta no es una definición completa porque las variables categóricas se refieren a cualquier tipo de variable que tome un número finito, posiblemente pequeño, de valores. Así, cualquier variable discreta y/o nominal y/u ordinal y/o cualitativa podría considerarse como una variable categórica.

Un caso muy concreto de variable agrupada o categórica es aquella variable que sólo toma dos valores, conocida como variable *binaria* o *dicotómica*.

En la Figura 12.1 se presenta un diagrama que facilita la comprensión entre las relaciones de los distintos tipos de variable mencionados. Los datos representados por variables cualitativas son siempre discretos; sin embargo, los datos representados por variables cuantitativas pueden ser tanto discretos (por ejemplo, talla de zapatos o una variable agrupada) como continuos (por ejemplo, temperatura). Las variables nominales son siempre cualitativas y discretas (por ejemplo, color de ojos), mientras que las variables continuas son siempre cuantitativas (por ejemplo, temperatura). Las variables categóricas pueden ser tanto cualitativas (por ejemplo, color de ojos) como cuantitativas (nivel de satisfacción en una escala del 1 al 5). Las variables categóricas nunca son continuas.

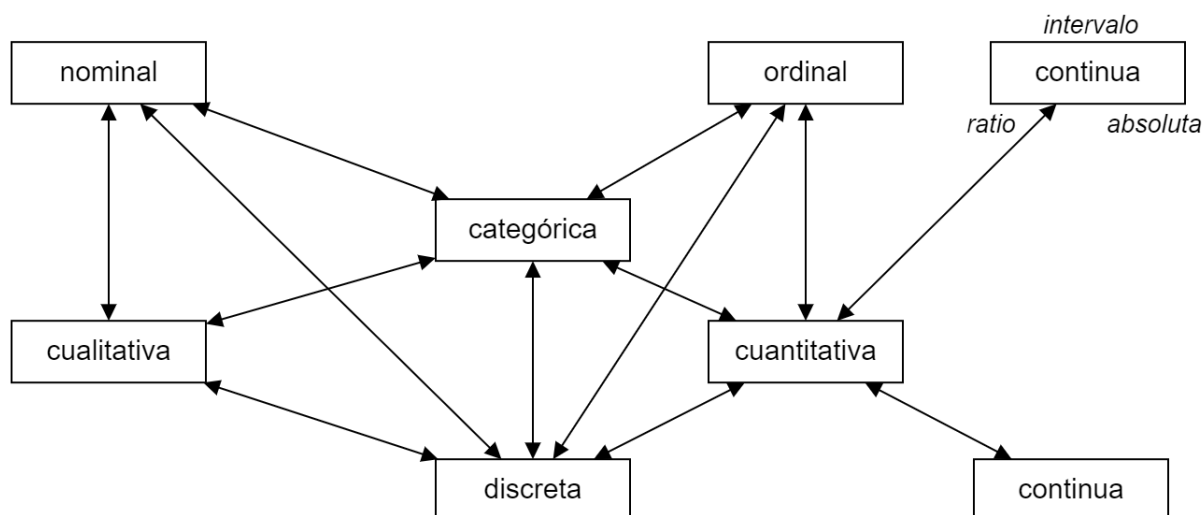


Figura 12.1: Diagrama. Relación entre los distintos tipos de variable.

12.4 Distribuciones unidimensionales de frecuencias

En las secciones previas se ha enfatizado cómo los distintos tipos de variables contienen diferentes niveles de información. Al resumir o visualizar una o más variables, la estructura de esta información es la que determina los métodos estadísticos apropiados para cada situación.

Supongamos que estamos interesados en estudiar las oportunidades de empleo y los salarios de los graduados universitarios. Se considera la variable X , que representa el salario inicial medido en €/año. Se conocen los salarios iniciales de 100 graduados, de manera que el salario inicial del primer estudiante será el valor x_1 , el del segundo estudiante x_2 , y así sucesivamente hasta el valor x_{100} , relativo al salario inicial del estudiante 100. Por lo tanto, hay 100 observaciones x_1, x_2, \dots, x_{100} . La cuestión es, ¿cómo se pueden resumir estos 100 valores de la mejor manera posible para extraer información significativa de ellos? La respuesta a esta pregunta depende de varios aspectos, como por ejemplo, la naturaleza de los datos registrados, la cantidad de observaciones que se han obtenido, cómo se registraron los datos (¿valores exactos o valores resumidos en intervalos?), por mencionar algunos matices. Por ejemplo, los salarios iniciales se pueden obtener como valores exactos, esto es 51500 €/año, 32350 €/año, etc. Alternativamente, estos valores se podrían resumir en categorías como por ejemplo, *salario bajo*, (< 30000 €/año); *salario medio*, ($30000 - 50000$ €/año); *salario alto*, ($50000 - 70000$ €/año); *salario muy alto*, (> 70000 €/año). Otro enfoque podría plantearse preguntando a los estudiantes si estaban empleados o no después de graduarse y registrar los datos en términos de *sí* o *no*. Es evidente que esta última clasificación es menos detallada que los datos sobre franjas salariales, que a su vez son menos detallados que los datos exactos. Dependiendo de qué conceptualización de *salario inicial* usemos, se debería elegir un enfoque u otro para resumir los datos, o lo que es lo mismo, los 100 valores relativos al salario inicial de cada uno de los 100 estudiantes graduados analizados.

12.4.1 Frecuencias absolutas y frecuencias relativas

- **Datos Discretos.** Se propone un ejemplo para ilustrar la notación.

Ejemplo 12.2:

Suponga que hay diez personas en la cola de un supermercado. Cada uno de ellos se codifica como F (si la persona es mujer) o M (si la persona es hombre). Los datos recopilados son:

$M, F, M, F, M, M, M, F, M, M$

Se puede observar que hay dos categorías en los datos: masculino (M) y femenino (F). Como hay 7 hombres y 3 mujeres, hay un total de 7 valores en la categoría M , denotados como $n_1 = 7$, y 3 valores en la categoría F , denotados como $n_2 = 3$. El número de observaciones en una categoría concreta se llama **frecuencia absoluta**, de manera que $n_1 = 7$ y $n_2 = 3$ son las frecuencias absolutas de M y F , respectivamente. Es importante tener en cuenta que, siendo n el número total de observaciones recopiladas, se cumple que $n_1 + n_2 = n = 10$. Las frecuencias relativas de M y F también se pueden calcular como $f_1 = f(M) = \frac{n_1}{n} = \frac{7}{10} = 0,7 = 70\%$, y $f_2 = f(F) = \frac{n_2}{n} = \frac{3}{10} = 0,3 = 30\%$, respectivamente.

Las frecuencias relativas proporcionan información sobre la cantidad de veces que se repite cada categoría con respecto del número total de observaciones, o lo que es lo mismo, la proporción de cada categoría; en este ejemplo concreto, las proporciones de clientes masculinos y femeninos en la cola.

A continuación se presenta una generalización de estos conceptos a un marco más amplio relativo al resumen de datos en variables discretas. Se supone la existencia de k categorías denotadas, como a_1, a_2, \dots, a_k , siendo n_j la cantidad de observaciones en la categoría j , para cada $j = 1, 2, \dots, k$. La **frecuencia absoluta** n_j se define como el número de unidades en la j -ésima categoría, a_j . Además, se cumple que la suma de frecuencias absolutas es igual al número total de unidades en los datos,

$$\sum_{j=1}^k n_j = n$$

donde, para cada $j = 1, 2, \dots, k$, la **frecuencia relativa** de la clase j se define como

$$f_j = f(a_j) = \frac{n_j}{n}, \quad j = 1, 2, \dots, k \quad (12.1)$$

Además, se cumple que

- $0 \leq f_j \leq 1$, para cada $j = 1, 2, \dots, k$.
- $\sum_{j=1}^k f_j = 1$

• Datos Continuos Agrupados.

Los datos sobre variables continuas suelen tener una gran cantidad (k) de valores diferentes. Incluso podría darse el caso de que k sea igual a n (número total de observaciones). en tal caso, cada una de las frecuencias relativas vale $f_j = \frac{1}{n}, \forall j = 1, 2, \dots, n$. Así, pese a que pueda existir una amplia gama de valores diferentes, es posible definir intervalos en los que están contenidos los valores observados.

Ejemplo 12.3:

Considere los siguientes $n = 20$ resultados de un examen, cuyo valor máximo son 100 puntos.

28, 35, 42, 90, 70, 56, 75, 66, 30, 89, 75, 64, 81, 69, 55, 83, 72, 68, 73, 16.

Estos resultados se pueden resumir en intervalos de clase del tipo $0 - 20$, $21 - 40$, $41 - 60$, $61 - 80$, y $81 - 100$, de manera que los datos se podrían presentar como:

| Intervalos de clase | Frecuencias Absolutas | Frecuencias Relativas |
|---------------------|-----------------------|-----------------------|
| 0-20 | $n_1 = 1$ | $f_1 = \frac{1}{20}$ |
| 21-40 | $n_2 = 3$ | $f_2 = \frac{3}{20}$ |
| 41-60 | $n_3 = 3$ | $f_3 = \frac{3}{20}$ |
| 61-80 | $n_4 = 9$ | $f_4 = \frac{9}{20}$ |
| 81-100 | $n_5 = 4$ | $f_5 = \frac{4}{20}$ |

Tabla 12.1: Distribución de frecuencias. Ejemplo 12.3.

Se puede observar que $\sum_{j=1}^5 n_j = 20$ y $\sum_{j=1}^5 f_j = 1$.

En un contexto más general, se asume que las n observaciones se pueden agrupar en k intervalos, a_1, a_2, \dots, a_k , de manera que, para cada $j = 1, 2, \dots, k$, el intervalo j -ésimo, a_j , contiene n_j observaciones, de manera que $\sum_{j=1}^k n_j = n$. La frecuencia relativa de la clase j -ésima es $f_j = \frac{n_j}{n}$, y además, $\sum_{j=1}^k f_j = 1$. En la Tabla 12.2 se presenta la **distribución de frecuencias** de una variable discreta, X .

| Intervalos de clase | Frecuencias Absolutas | Frecuencias Relativas |
|---------------------|-----------------------|-----------------------|
| a_1 | n_1 | f_1 |
| a_2 | n_2 | f_2 |
| \dots | \dots | \dots |
| a_k | n_k | f_k |

Tabla 12.2: Distribución de frecuencia de una variable discreta, X .**Ejemplo 12.4:**

Se consideran los datos de un servicio de entrega de comida a domicilio. En concreto, hay interés en el servicio de entrega por zonas, de manera que se genera una tabla de frecuencias considerando las zonas *Centro*, *Norte* y *Sur*, que muestra la distribución de los datos.

Se sabe que $n = \sum_j n_j = 1266$ entregas, y $\sum_j f_j = 1$. En la Tabla 12.3 se puede observar que todas las zonas tienen una frecuencia absoluta de envíos similar, y que la frecuencia relativa de los envíos de cada zona se corresponde aproximadamente con $\frac{1}{3}$ del total de envíos.

| a_j | n_j | f_j |
|---------------|-------|----------------------------|
| <i>Centro</i> | 421 | $\frac{421}{1266} = 0,333$ |
| <i>Norte</i> | 410 | $\frac{410}{1266} = 0,323$ |
| <i>Sur</i> | 435 | $\frac{435}{1266} = 0,344$ |

Tabla 12.3: Tabla de frecuencias. Ejemplo 12.4

12.5 Distribuciones acumuladas

Otra manera de resumir y visualizar la distribución (frecuencia) de las distintas variables es mediante la función de distribución acumulativa empírica. Como sugiere el propio nombre, esta función proporciona una idea sobre la acumulación relativa de frecuencias hasta cierto punto. Volviendo al Ejemplo 12.3, se quiere conocer la cantidad de personas que obtuvieron menos 60 puntos en el examen. Este resultado se puede calcular sumando el número de personas en los intervalos de clase $0 - 20$, $21 - 40$ y $41 - 60$, lo que se corresponde con $n_1 + n_2 + n_3 = 1 + 3 + 3 = 7$. Este valor se conoce como **frecuencia acumulada**. Si se quisiera conocer la frecuencia relativa de personas que obtienen hasta 60 puntos, habría que sumar las frecuencias relativas correspondientes a los intervalos $0 - 20$, $21 - 40$ y $41 - 60$, esto es, $f_1 + f_2 + f_3 = \frac{1}{20} + \frac{3}{20} + \frac{3}{20} = \frac{7}{20}$.

Antes de iniciar el discurso acerca de la función de distribución acumulativa empírica en un contexto más general, se introduce el concepto de **valores ordenados**. Para ello se propone el siguiente ejemplo.

Ejemplo 12.5:

Se considera una variable X que representa la altura de una persona. Se han observado cuatro personas, de manera que sus alturas se representan mediante los valores $x_1 = 180cm$, $x_2 = 160cm$, $x_3 = 175cm$, y $x_4 = 170cm$. Debido a la naturaleza de la variable X , estos valores se pueden organizar atendiendo a un orden, en concreto, el orden ascendente. De esta manera, el primer valor de la ordenación será el valor más pequeño de entre los cuatro disponibles, y se denotará como $x_{(1)}$; el segundo valor de la ordenación será el segundo valor más pequeño de entre los disponibles, y se denota como $x_{(2)}$, y así sucesivamente hasta llegar al valor más alto de los disponibles, que ocupará el último lugar de la ordenación, en este caso, $x_{(4)}$.

Entonces, se tiene:

- $x_{(1)} = 160cm = x_2$
- $x_{(2)} = 170cm = x_4$
- $x_{(3)} = 175cm = x_3$
- $x_{(4)} = 180cm = x_1$

Los valores $x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}$ se dice que son **valores ordenados**, para los cuales se cumple que $x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq x_{(4)}$. Es importante entender la diferencia entre los valores x_i y $x_{(i)}$. Así, el valor x_1 no tiene por qué ser necesariamente el valor más pequeño, a diferencia de $x_{(1)}$, que es

por definición el valor más bajo de todos los disponibles (o el más alto en caso de que se hubiera elegido un orden descendente).

En general, si hay n observaciones x_1, x_2, \dots, x_n , los datos ordenados son $x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(n)}$.

Definición 2.

Se consideran n observaciones de una variable X , $x_1, x_2, x_3, \dots, x_n$, ordenadas de forma ascendente de manera que $x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(n)}$ (y por lo tanto están en una escala al menos ordinal). La **función de distribución acumulativa empírica**, también conocida como **función de distribución acumulada**, $F(x)$, se define como el acumulado de todas las frecuencias relativas correspondientes a todos los valores a_j que son inferiores o iguales a x :

$$F(x) = \sum_{a_j \leq x} f(a_j) \quad (12.2)$$

Por su propia definición, una función de distribución acumulada tiene las siguientes propiedades (Pinto y Castillo Galarza 2017):

- Es monótona no decreciente.
- $0 \leq F(x)$, $\forall x \in \mathbb{R}$.
- $F(x) \leq 1$, $\forall x \in \mathbb{R}$.
- $\lim_{x \rightarrow -\infty} F(x) = 0$ (el límite inferior de F es 0).
- $\lim_{x \rightarrow +\infty} F(x) = 1$ (el límite superior de F es 1).

12.5.1 Función de distribución para variables ordinales

La función de distribución de toda variable ordinal es una función escalonada.

Ejemplo 12.6:

Se considera una encuesta de satisfacción del cliente llevada a cabo por un taller de automóviles. A los 200 clientes que han llevado a reparar un automóvil en los últimos 30 días se les pidió que contestaran a una pregunta relativa a su nivel general de satisfacción con la calidad del servicio, utilizando para ello una escala de 1 a 5 basada en las siguientes opciones: 1 = *no satisfecho en absoluto*, 2 = *insatisfecho*, 3 = *satisfecho*, 4 = *muy satisfecho* y 5 = *sumamente satisfecho*. Atendiendo a la frecuencia con que se repite cada opción, se pueden calcular las correspondientes frecuencias relativas, para después trazar la función de distribución acumulativa empírica, donde, para cada $j = 1, 2, 3, 4, 5$, la función de distribución correspondiente al valor a_j se calcula como

$$F(a_j) = F_j = \sum_{\ell=1}^j f_{\ell}.$$

| Nivel de satisfacción a_j | n_j | f_j | F_j |
|--------------------------------|-------|-------------------------|--|
| $j=1$ | 4 | $\frac{4}{200} = 0,02$ | $\frac{4}{200} = 0,02$ |
| $j=2$ | 16 | $\frac{16}{200} = 0,08$ | $\frac{4}{200} + \frac{16}{200} = 0,02 + 0,08 = 0,1$ |
| $j=3$ | 90 | $\frac{90}{200} = 0,45$ | $\frac{4}{200} + \frac{16}{200} + \frac{90}{200} = 0,02 + 0,08 + 0,45 = 0,55$ |
| $j=4$ | 70 | $\frac{70}{200} = 0,35$ | $\frac{4}{200} + \frac{16}{200} + \frac{90}{200} + \frac{70}{200} = 0,02 + 0,08 + 0,45 = 0,9$ |
| $j=5$ | 20 | $\frac{20}{200} = 0,1$ | $\frac{4}{200} + \frac{16}{200} + \frac{90}{200} + \frac{70}{200} + \frac{20}{200} = 0,02 + 0,08 + 0,45 + 0,1 = 1$ |

Tabla 12.4: Tabla de distribuciones. Ejercicio 12.6.

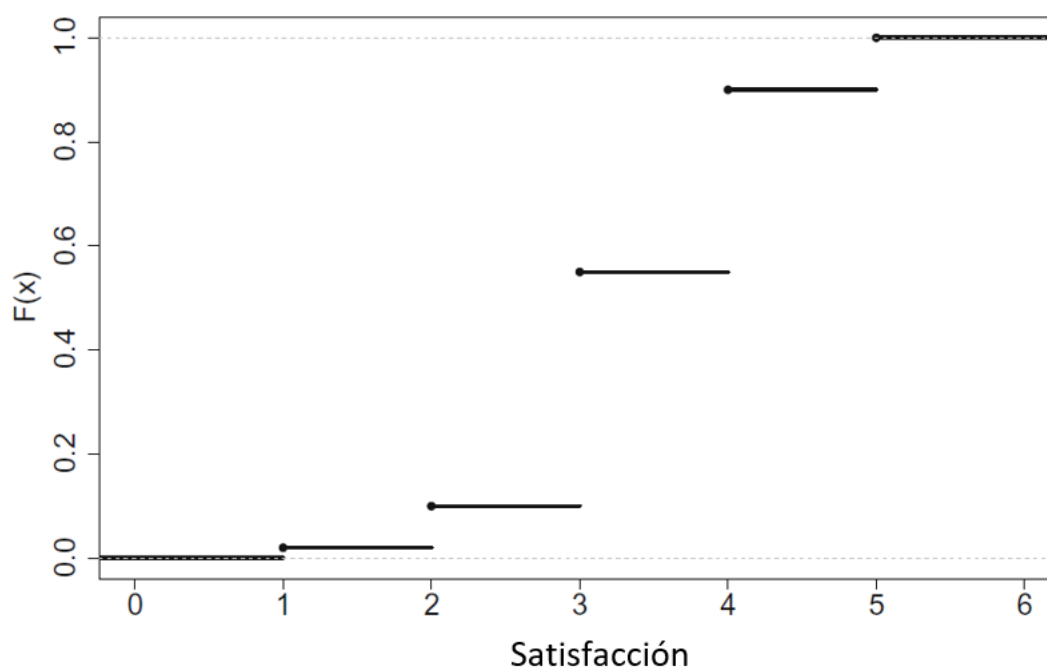


Figura 12.2: Función de distribución. Ejemplo 12.6

Las funciones de distribución acumuladas se pueden utilizar para obtener las frecuencias relativas de los valores contenidos en ciertos intervalos como

$$H(c \leq x \leq d) = \text{frecuencia relativa de los valores } x \text{ que cumplen } c \leq x \leq d$$

De donde se obtienen las siguientes propiedades:

$$H(x \leq a_j) = F(a_j) \quad (12.3)$$

$$H(x < a_j) = H(x \leq a_j) - f(a_j) = F(a_j) - f(a_j) \quad (12.4)$$

$$H(x > a_j) = 1 - H(x \leq a_j) = 1 - F(a_j) \quad (12.5)$$

$$H(x \geq a_j) = 1 - H(x < a_j) = 1 - F(a_j) + f(a_j) \quad (12.6)$$

$$H(a_{j_1} \leq x \leq a_{j_2}) = F(a_{j_2}) - F(a_{j_1}) + f(a_{j_1}) \quad (12.7)$$

$$H(a_{j_1} < x \leq a_{j_2}) = F(a_{j_2}) - F(a_{j_1}) \quad (12.8)$$

$$H(a_{j_1} < x < a_{j_2}) = F(a_{j_2}) - F(a_{j_1}) - f(a_{j_2}) \quad (12.9)$$

$$H(a_{j_1} \leq x < a_{j_2}) = F(a_{j_2}) - F(a_{j_1}) - f(a_{j_2}) + f(a_{j_1}) \quad (12.10)$$

Ejemplo 12.7

Retomando el Ejemplo 12.6, se supone que se quiere saber cuántos clientes no están satisfechos con el servicio recibido. Utilizando los datos relacionados con las respuestas “1” (*no satisfecho en absoluto*) y “2” (*insatisfecho*), a partir de la función de distribución, se puede observar que $\frac{16+4}{200} = 10\%$ de los clientes no estaban satisfechos con el servicio recibido. Esto se relaciona con el uso de la regla (12.3):

$$H(x \leq 2) = F(2) = 0,1 = 10\%$$

.

Del mismo modo, la proporción de clientes que están más que satisfechos se puede obtener usando (12.5) ya que

$$H(x > 3) = 1 - H(x \leq 3) = 1 - \frac{110}{200} = 0,45 = 45\%$$

12.5.2 Función de distribución para variables continuas

En general, las fórmulas (12.2) a (12.10) también se pueden aplicar a datos representados mediante variables continuas. Antes de mostrar su uso, se considera un escenario algo diferente. Se supone la existencia de una variable continua, cuya información de interés sólo está disponible de manera agrupada. Se puede suponer que las observaciones dentro de cada grupo (cada categoría o intervalo) se distribuyen uniformemente a lo largo de todo el grupo. En este escenario, la función de distribución consiste en líneas rectas que conectan los valores superior e inferior de la distribución correspondiente en cada intervalo. Para comprender este concepto con más detalle, se propone la siguiente notación:

- k es el número de grupos, intervalos o categorías
- e_{j-1} es el límite inferior del intervalo j -ésimo
- e_j es el límite superior del intervalo j -ésimo
- $d_j = e_j - e_{j-1}$ es la longitud del intervalo j -ésimo

- n_j es el número de observaciones del intervalo j -ésimo

Bajo el supuesto de que todos los valores en un intervalo particular se distribuyen uniformemente dentro de dicho intervalo, la función de distribución acumulada se relaciona con una cadena poligonal que conecta los puntos $(0,0), (e_1, F(e_1)), (e_2, F(e_2)), \dots, (e_k, 1)$. Así, siendo $F(e_0) = 0$, la función de distribución acumulada se puede definir como

$$F(x) = \begin{cases} 0, & x < e_0 \\ F(e_j) + \frac{f_j}{d_j}(x - e_{j-1}), & x \in [e_{j-1}, e_j] \\ 1, & x \geq e_k \end{cases} \quad (12.11)$$

La idea representada en la expresión (12.11) se puede observar en la Figura 12.3. Para cualquier intervalo $[e_{j-1}, e_j]$, los límites inferior y superior relativos a la función de distribución son $F(e_j)$ y $F(e_{j-1})$. Si se asume que los valores se distribuyen uniformemente en este intervalo, los valores $F(e_j)$ y $F(e_{j-1})$ se pueden conectar mediante una línea recta. Para obtener $F(x)$, siendo $x > e_{j-1}$ y $x < e_j$, simplemente se agrega la altura entre los puntos $F(e_{j-1})$ y $F(x)$ hasta $F(e_j)$.

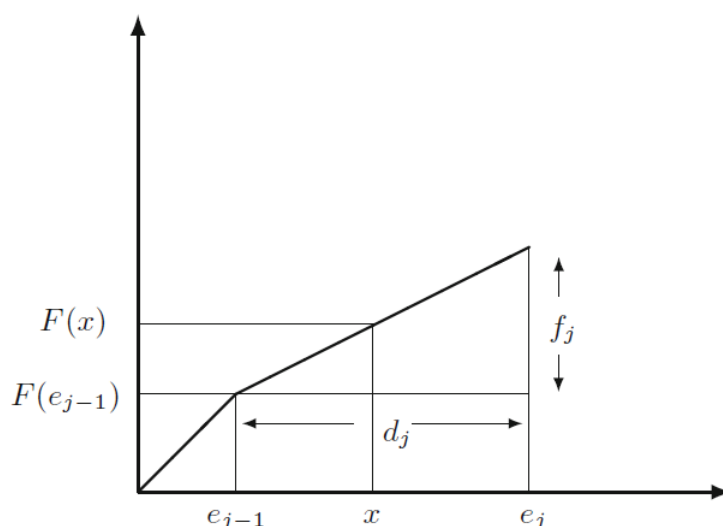


Figura 12.3: Función de distribución acumulada para datos representados mediante variables continuas cuando se tienen en formato acumulado.

Ejemplo 12.8:

Retomemos el Ejemplo 12.4 relativo a un servicio de entrega de comida a domicilio. Pese a que la estructura de la curva es una función escalonada, en este ejemplo casi parece una curva continua. Esto se debe a que, cuando el número de observaciones es grande, la duración de los intervalos de clase se vuelve pequeña. Cuando estas pequeñas longitudes se unen, aparecen como una curva continua. Cuando el número de observaciones aumenta, la suavidad de la curva también aumenta. Si no hubiera muchas observaciones, la función de distribución se podría caracterizar con una tabla que resuma todas las características, como se muestra en la Tabla

12.5. Es interesante ver que los gráficos que surgen del uso de los datos agrupados y los datos no agrupados son similares en este ejemplo específico, como se puede observar en la Figura 12.4.

Supongamos que estamos interesados en calcular cuántas entregas se completaron dentro del límite de tiempo deseado de 30 minutos, con una tolerancia máxima del 10 % de desviación, esto es, 3 minutos. Se puede evaluar la función de distribución en $x = 3$ min.

Atendiendo a la ecuación (12.11), se calcula $H(x \leq 33) = F(33) = F(30) + \frac{f(6)}{5}(33 - 30) = 0,248 + \frac{0,2946}{5}3 = 0,4248$. Así, en base a los datos agregados, se concluye que sólo el 42 % de las entregas fueron completadas en la franja horaria deseada.

| Intervalo de entrega | j | e_{j-1} | e_j | n_j | f_j | $F(e_j)$ |
|----------------------|----|-----------|-------|-------|--------|----------|
| [0; 10] | 1 | 0 | 10 | 0 | 0 | 0 |
| (10; 15] | 2 | 10 | 15 | 3 | 0.0024 | 0.0024 |
| (15; 20] | 3 | 15 | 20 | 21 | 0.0166 | 0.019 |
| (20; 25] | 4 | 20 | 25 | 75 | 0.0592 | 0.0782 |
| (25; 30] | 5 | 25 | 30 | 215 | 0.1698 | 0.248 |
| (30; 35] | 6 | 30 | 35 | 373 | 0.2946 | 0.5426 |
| (35; 40] | 7 | 35 | 40 | 350 | 0.2765 | 0.8191 |
| (40; 45] | 8 | 40 | 45 | 171 | 0.1351 | 0.9542 |
| (45; 50] | 9 | 45 | 50 | 52 | 0.0411 | 0.9953 |
| (50; 55] | 10 | 50 | 55 | 6 | 0.0047 | 1 |

Tabla 12.5: Valores para calcular la función de distribución del Ejemplo 12.4.

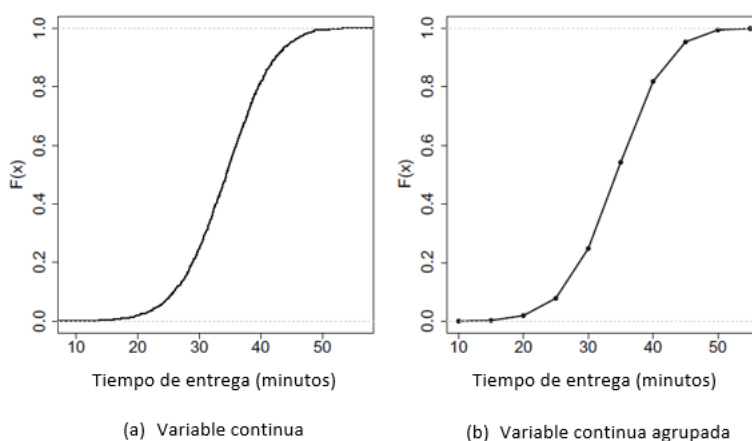


Figura 12.4: Función de distribución para datos continuos y continuos agrupados.

12.6 Tablas estadísticas

En teoría, las funciones estadísticas, tanto continuas como discretas, permiten determinar las probabilidades de un suceso, partiendo del modelo estadístico al que ese suceso se ajusta.

En la práctica, si se quiere conocer el valor numérico de dicha probabilidad, además de la expresión que la determina, es necesario cuantificar la distribución de probabilidad. Estas expresiones no son en absoluto sencillas; para facilitar su manipulación e interpretación se propone la utilización de tablas estadísticas. Una tabla estadística es una tabla en la que se dispone de forma ordenada y agrupada de los valores y frecuencias de una distribución. Es importante diferenciar entre tablas estadísticas de distribuciones no agrupadas y tablas de distribuciones agrupadas.

Las tablas de distribuciones no agrupadas contienen la siguiente información: los valores de la distribución, ordenados de menor a mayor si son caracteres cuantitativos; y las frecuencias relativas. En caso de que se puedan definir las frecuencias acumuladas, éstas también se añaden en términos absoluto y relativo.

Por su parte, las tablas estadísticas de distribuciones de frecuencia agrupadas por intervalos, contienen la siguiente información: los intervalos considerados; las amplitudes de los intervalos; las marcas de clase, (que son puntos medios de los intervalos); las frecuencias absolutas de cada intervalo; las frecuencias relativas. También se suelen incorporar las frecuencias acumuladas. Si todos los intervalos considerados fueran de la misma amplitud, se obvia la información sobre la amplitud.

12.7 Representación gráfica

Las tablas de frecuencia y las funciones de distribución acumulativa empírica son útiles para proporcionar un resumen numérico de una variable. Una forma alternativa de resumir la información de una variable es mediante el uso de gráficos. En muchas situaciones, los gráficos tienen la ventaja de transmitir la información oculta en los datos de forma más compacta. Algunos de los tipos más populares de gráficos se presentan a continuación.

12.7.1 Gráfico de barras

Los gráficos de barras son una herramienta muy sencilla para visualizar las frecuencias relativas o absolutas de los valores observados de una variable. Los gráficos de barras se pueden utilizar tanto para variables nominales como ordinales, siempre que el número de categorías no sea muy grande. Así, un gráfico de barras tendrá una barra para cada categoría. La altura de cada barra está determinada por la frecuencia absoluta o la frecuencia relativa de la categoría correspondiente; esta altura se muestra en el eje Y . Si la variable de interés fuera ordinal, se recomienda organizar las barras en el eje X atendiendo a sus rangos o valores. Si el número de categorías es muy grande, tal vez no sea conveniente el uso de este tipo de gráficos, ya que la cantidad de barras aumentará.

Cabe mencionar que las barras usadas pueden ser horizontales en vez de verticales.

Ejemplo 12.9:

Se considera el Ejemplo 12.2 sobre un grupo de diez personas que hacen cola en un supermercado. Estas personas están clasificadas como hombres (M) o mujeres (F). Las frecuencias absolutas para estas categorías son $n_1 = 7$ y $n_2 = 3$, respectivamente. Dado que hay dos categorías, M y F , harán falta dos barras para construir el gráfico, una para cada categoría. Las alturas de las

barras se determinan como $n_1 = 7$ y $n_2 = 3$ o $f_1 = 0,7$ y $f_2 = 0,3$. Estos gráficos se muestran en la Figura 12.5.

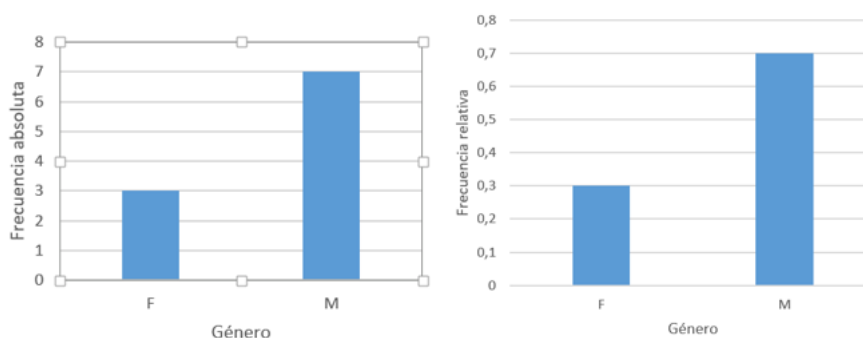


Figura 12.5: Gráficos de barras relativos al Ejemplo 12.2, con frecuencias absolutas y relativas.

Ejemplo 12.10:

Se consideran los datos del Ejemplo 12.8, acerca de los tiempos de un servicio de entrega de comida a domicilio, que se presentan en la Tabla 12.5. La tabla de frecuencias forma la base para el gráfico de barras, ya sea utilizando las frecuencias absolutas o relativas en el eje Y . La Figura 12.6 muestra los gráficos de barras para el número y la proporción de entregas.

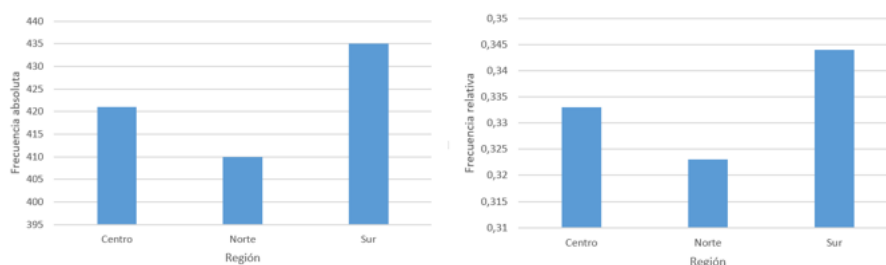


Figura 12.6: Gráficos de barras relativos al Ejemplo 12.8, con frecuencias absolutas y relativas.

12.7.2 Gráfico de tarta

El gráfico de tarta, también conocido como gráfico circular o de sectores, es otra opción para visualizar las frecuencias absolutas y relativas de variables nominales y ordinales. Un gráfico circular es un círculo dividido en sectores, donde cada uno de los sectores representa una categoría. El tamaño de cada sector depende de la frecuencia relativa de la categoría correspondiente, y está determinada por el ángulo $f_j * 360^\circ$, para la categoría a_j .

Ejemplo 12.11:

Para ilustrar la construcción de un gráfico de tarta, se consideran una vez más los datos del

Ejemplo 12.2 sobre un grupo de diez personas que hacen cola en un supermercado, clasificadas como $M, F, M, F, M, M, M, F, M, M$. Para estos datos, el gráfico circular tendrá dos sectores, uno para representar a los hombres y otro para representar a las mujeres. Las frecuencias relativas correspondientes son $f_1 = \frac{7}{10}$ y $f_2 = \frac{3}{10}$, respectivamente. Por tanto, el tamaño del sector que representa a los hombres (M) será $f_1 * 360^\circ = \frac{7}{10} * 360^\circ = 252^\circ$, mientras que el tamaño del sector que representa a las mujeres (F) será $f_2 * 360^\circ = \frac{3}{10} * 360^\circ = 108^\circ$. El gráfico de tarta obtenido se puede observar en la Figura 12.7.

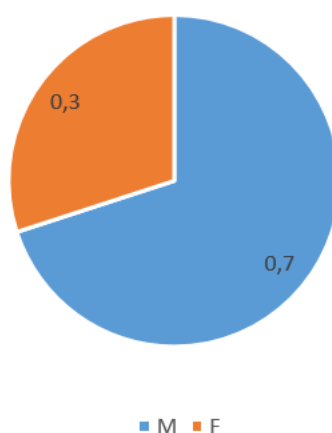


Figura 12.7: Gráficos de tarta relativo al Ejemplo 12.2.

Ejemplo 12.12:

Se consideran los datos del Ejemplo 12.6 relativos a la satisfacción de 200 clientes con un servicio, clasificada en cinco categorías que comprenden las opiniones *no satisfecho en absoluto* a *sumamente satisfecho*. En este caso, el gráfico de tarta constará de cinco sectores, cada uno de los cuales representará una de las categorías de opinión, 1, 2, 3, 4 y 5. El tamaño del sector j -ésimo es $f_j * 360^\circ$. Por ejemplo, 4 de cada 200 clientes están *no satisfechos en absoluto*, con lo cual pertenecen a la categoría 1. En este caso, el ángulo del sector correspondiente es $f_1 * 360^\circ = \frac{4}{200} * 360^\circ = 7,2^\circ$. El resto de sectores se calculan de manera similar. El gráfico de tarta obtenido se puede observar en la Figura 12.8.



Figura 12.8: Gráficos de tarta relativo al Ejemplo 12.6.

A la hora de construir un gráfico de tarta, es importante tener en cuenta que el área de cada sector no es proporcional al valor absoluto de la frecuencia absoluta, sino proporcional al ángulo $f_j * 360^\circ$. Este hecho podría causar interpretaciones incorrectas ya que el ojo humano puede captar el área de un sector más fácilmente que el ángulo de un sector.

12.7.3 Histograma

Si una variable tiene un número alto de valores diferentes, la cantidad de categorías utilizadas para construir gráficos de barras también será grande. Por tanto, cuando se trata de representar variables continuas, los gráficos de barras podrían no ser una buena opción. En este escenario, la opción adecuada es la representación mediante un histograma, que permite mostrar la distribución de valores de variables continuas. La construcción de un histograma se basa en la idea de categorizar los datos en diferentes grupos, y después trazar las barras para cada categoría con altura $h_j = \frac{f_j}{d_j}$, donde $d_j = e_j - e_{j-1}$ denota el ancho del j -ésimo intervalo o categoría. Una consideración importante para este concepto es que el área de las barras (alto \times ancho, o base \times altura) es proporcional a la frecuencia relativa. Esto significa que los anchos de todas las barras no tienen por qué ser necesariamente los mismos en un histograma; este tamaño podría ajustarse con la altura para así obtener el área adecuada.

Ejemplo 12.13:

Se consideran los datos del Ejemplo 12.3 relativos a la puntuación obtenida por un grupo de $n = 20$ personas en un examen. Las puntuaciones se clasifican en los intervalos $0 - 20$, $21 - 40$, $41 - 60$, $61 - 80$, $81 - 100$, de manera que la tabla de frecuencias correspondientes es:

| Intervalo de clase | Frecuencia absoluta | Frecuencia relativa | Altura $h_j = \frac{f_j}{d_j}$ |
|--------------------|---------------------|----------------------|-----------------------------------|
| 0-20 | $n_1 = 1$ | $f_1 = \frac{1}{20}$ | $h_1 = \frac{1}{400}$ |
| 21-40 | $n_2 = 3$ | $f_2 = \frac{3}{20}$ | $h_2 = \frac{3}{400}$ |
| 41-60 | $n_3 = 3$ | $f_3 = \frac{3}{20}$ | $h_3 = \frac{3}{400}$ |
| 61-80 | $n_4 = 3$ | $f_4 = \frac{9}{20}$ | $h_4 = \frac{9}{400}$ |
| 81-100 | $n_5 = 4$ | $f_5 = \frac{4}{20}$ | $h_5 = \frac{4}{400}$ |

Tabla 12.6: Valores para calcular el histograma relativo al Ejemplo 12.3

El histograma correspondiente a estos datos se puede ver en la Figura 12.9. Como los datos se han agrupado en 5 categorías, el histograma consta de 5 barras. Puesto que los anchos de los intervalos de clase son los mismos, las alturas de las barras son proporcionales a la frecuencia relativa de la categoría respectiva

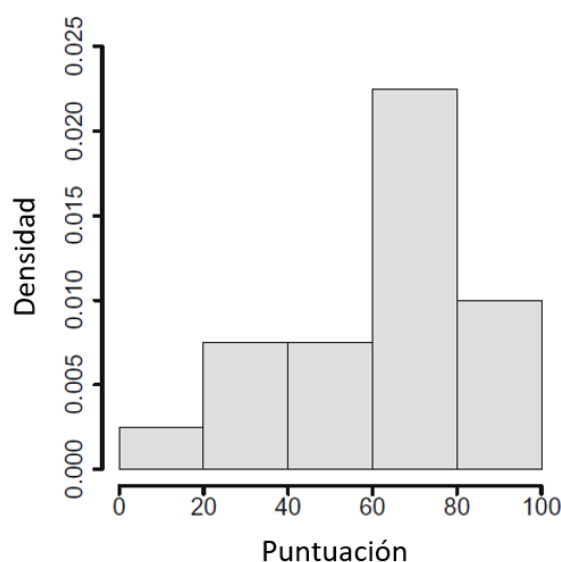


Figura 12.9: Histograma relativo al Ejemplo 12.3.

Ejemplo 12.4:

Se retoma el Ejemplo 12.4 relativo a un servicio de entrega de comida a domicilio. En la Tabla 12.7 se muestran los datos necesarios para construir un histograma que represente esta situación. En el gráfico de la izquierda de la Figura 12.10 se muestra el histograma con anchos iguales de intervalos de tiempo de entrega. Se puede ver una distribución simétrica de los tiempos de entrega; no obstante, algunas entregas superan el tiempo objetivo de 30 minutos. Si se requiere que el histograma tenga diferentes anchos para diferentes barras, es decir, diferentes intervalos de tiempo de entrega para diferentes categorías, se podría construir como se muestra en el gráfico de la derecha de la Figura 12.10.

| Intervalo de entrega | j | e_{j-1} | e_j | n_j | f_j | h_j |
|----------------------|----|-----------|-------|-------|--------|---------|
| [0; 10] | 1 | 0 | 10 | 0 | 0 | 0.00000 |
| (10; 15] | 2 | 10 | 15 | 3 | 0.0024 | 0.00047 |
| (15; 20] | 3 | 15 | 20 | 21 | 0.0166 | 0.00332 |
| (20; 25] | 4 | 20 | 25 | 75 | 0.0592 | 0.01185 |
| (25; 30] | 5 | 25 | 30 | 215 | 0.1698 | 0.03397 |
| (30; 35] | 6 | 30 | 35 | 373 | 0.2946 | 0.05893 |
| (35; 40] | 7 | 35 | 40 | 350 | 0.2765 | 0.05529 |
| (40; 45] | 8 | 40 | 45 | 171 | 0.1351 | 0.02701 |
| (45; 50] | 9 | 45 | 50 | 52 | 0.0411 | 0.00821 |
| (50; 55] | 10 | 50 | 55 | 6 | 0.0047 | 0.00094 |

Tabla 12.7: Valores para calcular la función de distribución del Ejemplo 12.4.

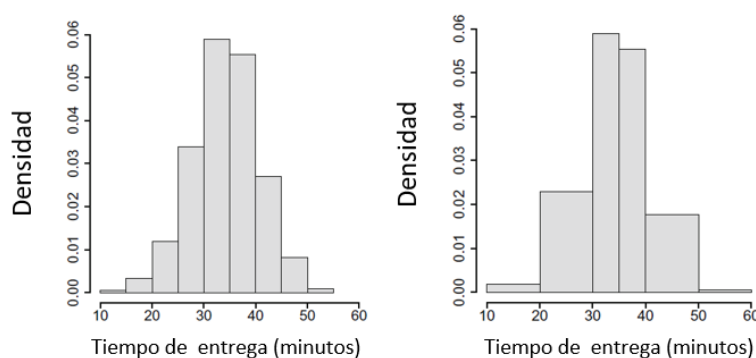


Figura 12.10: Histograma relativo al Ejemplo 12.4.

12.7.4 Estimación Kernel de la densidad

Una desventaja de los histogramas es que los datos continuos se categorizan artificialmente. La elección de los intervalos de clase es crucial para el aspecto final del gráfico. Una forma más elegante de abordar este problema es suavizar el histograma en el sentido de que cada observación pueda contribuir a diferentes clases con diferentes pesos. Así, la distribución estaría representada por una función continua en lugar de una función escalonada. Una estimación Kernel de la densidad se puede conseguir con la función:

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right), \quad h > 0 \quad (12.12)$$

donde $n > 0$ es el tamaño de la muestra, h el ancho de cada barra y K la función Kernel. Algunos ejemplos de función Kernel son la rectangular (K_1) y la Epanechnikov (K_2):

$$K_1(x) = \begin{cases} \frac{1}{2}, & -1 < x \leq 1 \\ 0, & \text{en otro caso} \end{cases} \quad (12.13)$$

$$K_2(x) = \begin{cases} \frac{3}{4}(1-x^2), & |x| < 1 \\ 0, & \text{en otro caso} \end{cases} \quad (12.14)$$

Para comprender mejor este concepto, consideremos la Figura 12.11. Las marcas de graduación en el eje X representan cinco observaciones: 3, 6, 7, 8 y 10. En cada observación x_i así como en su valores circundantes, aplicamos una función Kernel, en concreto, el kernel de Epanechnikov. Esto significa que tenemos cinco funciones que se refieren a las cinco observaciones. Estas funciones son mayores en la propia observación y se convierten gradualmente en más pequeñas a medida que aumenta la distancia desde la observación. Agregando esta función Kernel tal y como se describe en la ecuación (12.14), se obtiene la línea negra continua. Es una curva suave, que representa la distribución de los datos. El grado de suavidad se puede controlar mediante el ancho h , que en este caso concreto es $h = 2$.

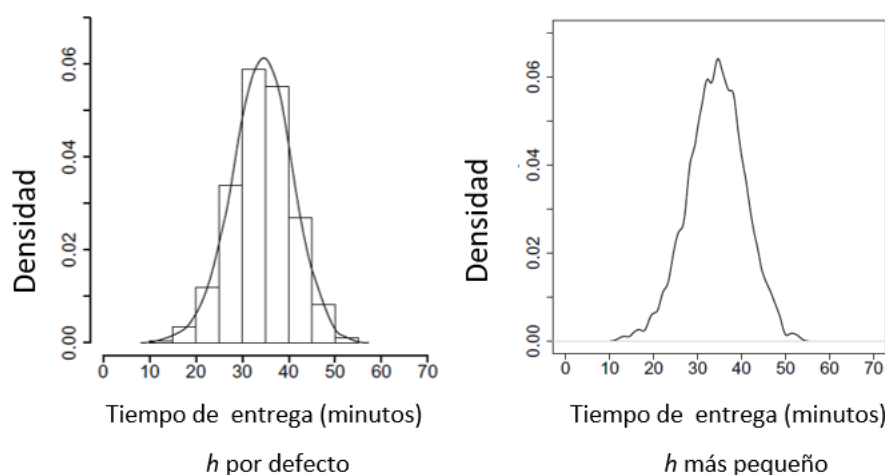


Figura 12.11: Estimación Kernel de la densidad para el tiempo de entrega

La elección del núcleo puede afectar el aspecto general de la gráfica. Antes se han propuesto las funciones para los núcleos rectangular y Epanechnikov. Otra de las funciones más comunes es la función de distribución normal. El núcleo que se basa en la distribución normal se denomina "núcleo gaussiano". Es importante tener en cuenta que las funciones del kernel no se definen arbitrariamente y deben satisfacer ciertas condiciones, como las requeridas para las funciones de densidad de probabilidad.

12.8 Algunas observaciones importantes

- Los gráficos de barras y los histogramas no son las mismas herramientas gráficas. Los gráficos de barras se usan para visualizar las categorías de variables nominales u ordinales, mientras que los histogramas permiten visualizar la distribución de variables continuas. Un gráfico de barras no requiere tener valores ordenados en el eje X , mientras que un histograma siempre requiere que los valores en el eje X estén en una escala continua. La interpretación de un histograma se simplifica si todos los intervalos de clase tienen

el mismo tamaño, ya que entonces las alturas de los rectángulos del histograma son proporcionales a las frecuencias absolutas o relativas.

- La función de distribución acumulada sólo se puede utilizar para variables continuas.
- Un gráfico de tarta resume las observaciones de una variable discreta (nominal, ordinal o variable continua agrupada). Solo es útil si la cantidad de valores diferentes (categorías) es pequeña. Debe tenerse en cuenta que el área de cada sector no es proporcional a la frecuencia absoluta de la categoría correspondiente; sin embargo, el ángulo del sector sí es proporcional a la frecuencia relativa de la categoría.
- Otras posibilidades para visualizar la distribución de variables son, por ejemplo, los diagramas de caja y los diagramas estratificados.
- Se han expuesto algunas de las representaciones más clásicas; hoy en día, el avance de las técnicas informáticas hace que con frecuencia los diagramas de barras y de sectores sean sustituidos por dibujos de tamaños proporcionales a las frecuencias, donde los dibujos son alusivos a la variable que se está representando.

Bibliografía

1. Rustom Jabbaz, A. (2012). *Estadística descriptiva, probabilidad e inferencia. Una visión conceptual y aplicada*. Santiago de Chile: Departamento de Economía Agraria, Facultad de Ciencias Agronómicas, Universidad de Chile.
2. Heumann, C. y Schomaker Shalabh, M. (2016). *Introduction to Statistics and Data Analysis, with Exercises, Solutions and Applications in R*. Springer.
3. Pinto, C.S. y del Castillo Galarza, S. (2017). *Fundamentos básicos de estadística*. Quito.

Tema 13

Estadística descriptiva II. Medidas de síntesis de una distribución de frecuencias. Medidas de posición. Media aritmética, geométrica y armónica. Cálculo de las mismas y propiedades. Aplicaciones.

13.1 Estadística descriptiva II

En general, toda variable se puede resumir de tres formas distintas. En el Capítulo 12 se proponen dos de ellas, a saber, el uso de tablas estadísticas, y la construcción de diferentes tipos de gráficos. En este capítulo se presentan una serie de *medidas resumen*, que se corresponden con parámetros o estadígrafos, dependiendo de si se trata con poblaciones o con muestras, y que sirven para representar el posicionamiento de los datos. Son las llamadas medidas de posición.

Los tres tipos mencionados de presentación de datos pueden ser elegidos de forma excluyente o complementaria, incluso los tres simultáneamente. No obstante, cada tipo de variable tiene sus propias particularidades.

13.2 Medidas de síntesis de una distribución de frecuencias

La frecuencia (absoluta) es el número de veces que una característica o valor se repite en un conjunto de datos. Una distribución de frecuencias es la presentación ordenada de los valores de una o más variables acompañados de sus respectivas frecuencias (absoluta y relativa). Las distribuciones o tablas de frecuencias permiten resumir los datos en una tabla que recoge:

- valores de la variable o modalidades del atributo,
- frecuencia absoluta o número de veces que aparece cada valor o modalidad en la muestra,
- porcentaje de veces que aparece cada valor de la variable o modalidad del atributo sobre el total de observaciones,
- porcentaje válido calculado sobre el total de observaciones excluidos los valores missing o valores perdidos,
- porcentaje acumulado hasta cada uno de los valores de la variable ordenados de menor a mayor. Este porcentaje tiene interpretación sólo en los casos en que la variable sea susceptible de medida por lo menos en una escala ordinal.

Atendiendo a la naturaleza de la variable, se puede diferenciar entre distribuciones de frecuencia relativas a:

- Datos categóricos: variables ordinales y nominales.
- Distribuciones de frecuencias numéricas: series simples y series agrupadas.

13.2.1 Distribución de frecuencias de variables cualitativas

Cuando se tratan variables en **escala nominal**, se recomienda presentar las frecuencias en orden decreciente.

Ejemplo 13.1.

En este ejemplo se dispone de una variable *Especialidad* relativa a los cargos concretos de ciertos sanitarios. Se conoce el número de especialistas de cada tipo de los considerados, representados mediante la frecuencia absoluta. En este caso, como se trata de una variable nominal, para facilitar la comprensión de la distribución de frecuencias, es conveniente presentarlas por orden descendiente, mostrando en primer lugar el valor (especialidad) que más ocurrencias tenga. En este caso, *Ocupación=Médico*.

| Ocupación | Frecuencia absoluta | Frecuencia relativa | Porcentaje |
|---------------|---------------------|---------------------|------------|
| Médico | 224 | 0.338 | 33.8 |
| Obstetra | 194 | 0.293 | 29.3 |
| Tecnólogo | 131 | 0.198 | 19.8 |
| Enfermero | 67 | 0.101 | 10.1 |
| Nutricionista | 47 | 0.0071 | 7.1 |
| Total | 663 | 1.00 | 100 |

Por otra parte, al tratar con variables en **escala ordinal**, lo recomendable es presentar las frecuencias en el orden de la jerarquía de la variable en cuestión.

Ejemplo 13.2.

En este ejemplo se considera una encuesta de satisfacción, cuyos participantes han respondido con las opciones *Muy buena - Buena - Regular - Mala - Muy mala*. Se dispone de la cantidad de respuestas de cada tipo, representada mediante la frecuencia absoluta. En este caso, como la variable *Opinión* es gradual, se puede ordenar, por lo que es conveniente presentar los datos atendiendo a este orden. En este caso, el primer valor presentado sería *Muy buena*.

| Opinión | Frecuencia absoluta | Frecuencia relativa | Porcentaje |
|-----------|---------------------|---------------------|------------|
| Muy buena | 35 | 0.179 | 17.9 |
| Buena | 73 | 0.372 | 37.2 |
| Regular | 58 | 0.296 | 29.6 |
| Mala | 23 | 0.117 | 11.7 |
| Muy mala | 7 | 0.036 | 3.6 |
| Total | 196 | 1.00 | 100 |

13.2.2 Distribución de frecuencias de variables cuantitativas

A la hora de tratar con variables cuantitativas, distinguiremos entre **series simples** y **series agrupadas**, cuando los datos están agrupados de alguna manera.

Cuando se trata de series simples, para cada valor de la variable se presenta su frecuencia. Esto es factible cuando no hay muchos valores distintos. En este escenario, las frecuencias deben presentarse en el orden de los valores de la variable. También podrían presentarse las frecuencias acumuladas.

Ejemplo 13.3.

En este ejemplo, se considera la cantidad de habitaciones que tienen ciertas viviendas. Se mide cuántas viviendas de cada tipo hay, entendiendo por tipo la cantidad de habitaciones. En la siguiente tabla, se presenta, para cada tipo de vivienda, la frecuencia absoluta, la frecuencia acumulada (por ejemplo, frecuencia de viviendas con 3 habitaciones o menos), y los respectivos porcentajes respecto del total de viviendas analizadas.

| Número de habitaciones | Frecuencia absoluta | Frecuencia abs. acumulada | Porcentaje | Porcentaje acumulado |
|------------------------|---------------------|---------------------------|------------|----------------------|
| 1 | 31 | 31 | 26.5 | 26.5 |
| 2 | 59 | 90 | 50.4 | 76.9 |
| 3 | 16 | 106 | 13.7 | 90.6 |
| 4 | 9 | 115 | 7.7 | 98.3 |
| 5 | 2 | 117 | 1.7 | 100.0 |
| TOTAL | 117 | | 100.0 | |

Ejemplo 13.4.

En este ejemplo se considera la cantidad de jóvenes de cada edad que forman parte de una asociación. En la siguiente tabla se presentan las distintas categorías y su frecuencia absoluta.

| Edad | Frecuencia absoluta | Porcentaje |
|------|---------------------|------------|
| 16 | 23 | 6 |
| 17 | 30 | 8 |
| 18 | 38 | 10 |
| 19 | 49 | 13 |
| 20 | 58 | 15 |
| 21 | 51 | 13 |
| 22 | 39 | 10 |
| 23 | 33 | 9 |
| 24 | 30 | 8 |
| 25 | 19 | 5 |
| 26 | 8 | 2 |
| 27 | 8 | 2 |

El ejemplo anterior es un caso de variable en la que no es conveniente mostrar todos los valores, por ser estos demasiados. Así, cuando hay múltiples valores diferentes en una variable cuantitativa, lo ideal es presentar una distribución agrupada de las frecuencias.

En escenarios de este tipo, los valores de una variable cuantitativa se agrupan en intervalos de clase. Para cada intervalo, se presenta su frecuencia, porcentaje, y en ocasiones, la frecuencia acumulada. En este caso, las frecuencias deben presentarse en el orden de los intervalos de clase.

Existen múltiples técnicas estadísticas para definir los intervalos de clase, aunque no resultan especialmente relevantes. Se recomienda definir los intervalos atendiendo a criterios que se ajusten a la realidad de la situación abordada.

Ejemplo 13.5.

En este caso, se presentan los datos del ejemplo anterior acerca de las edades de los jóvenes, de manera agrupada, considerando distintos intervalos de edad.

| Edad | Frecuencia absoluta | Porcentaje |
|-------|---------------------|------------|
| 16-19 | 140 | 36.3 |
| 20-24 | 211 | 54.7 |
| 25-27 | 35 | 9.0 |
| TOTAL | 386 | 100.0 |

13.3 Medidas de posición

Una tendencia humana natural es hacer comparaciones con el promedio. Por ejemplo, un estudiante que obtenga un 6 %, en un examen estará contento con el resultado si el promedio de la puntuación de la clase es del 3 %. Sin embargo, si la puntuación promedio de la clase es del 9 %, la nota de este estudiante “no será buena” en relación con la media, incluso aunque obtuviera un 7 %. Algunos otros ejemplos del uso del promedio son en el cálculo de la altura corporal media, la temperatura media en cierta ciudad, la carrera solicitada, la nota media en la Evau, el programa de televisión más popular de una franja horaria o el ingreso promedio de cierto grupo de trabajadores. Existen varios conceptos estadísticos que se refieren al “promedio” de los datos. La elección correcta de cuál de ellos considerar depende de la naturaleza y escala de los datos, así como del objetivo del estudio. Las funciones estadísticas que describen el promedio o la tendencia central se conocen como medidas de posición. Las medidas de posición son valores que permiten dividir el conjunto de datos en partes porcentuales iguales y se usan para clasificar una observación dentro de una población o muestra. Estos indicadores estadísticos permiten resumir los datos en uno solo, o dividir su distribución en intervalos del mismo tamaño. Se puede considerar, informalmente, que las medidas de posición sirven para “dividir” y “medir”. De esta forma, unos resumirán los diferentes valores en uno que, en este caso, sea representativo. Por ejemplo, un promedio. Mientras los otros dividirán el conjunto de los datos en partes iguales, más sencillas de interpretar; estaríamos hablando de los cuantiles.

El cálculo de las medidas de posición es una parte muy importante de cualquier estudio estadístico. Obtener estas medidas es el primer paso que se ha de dar en un análisis descriptivo.

Cuando se quiere conocer información sobre un fenómeno, hay que empezar haciendo una recopilación de los datos. Sin embargo, los datos por sí mismos, no aportan información relevante, lo que hace imprescindible un análisis de los mismos. Las medidas de posición, junto con las de dispersión, facilitan la agrupación y codificación de los datos.

Estas medidas constituyen el conocimiento principal y básico en estadística. Si no se afianzan conceptos como el promedio, es complicado entender nociones más complejas como lo son la regresión o el contraste de hipótesis.

En este tema se hablará de medidas de posición centrales, medidas que permiten resumir la distribución de los datos en un solo valor central, alrededor del cual se sitúan. En concreto, se introducen la media aritmética, la media geométrica y la media armónica. Son tres medidas centrales que nos indican un promedio ponderado de los datos. La primera es la más utilizada y la más conocida de las tres. La geométrica se aplica en series que muestran crecimientos porcentuales. Por su parte, la armónica es útil en el análisis de inversiones en bolsa.

En temas posteriores se hablará de otras medidas de posición central robustas, como la mediana, y la moda, y de medidas de posición no central, conocidas como cuantiles.

13.3.1 Media aritmética

Existen muchos tipos de media y la más conocida es la media aritmética. Sin embargo, la idea general se mantiene en todos ellos: conocer el valor “promedio”. Al ser una medida de tendencia central, lo que se pretende mediante el cálculo de la media aritmética es aportar información sobre el centro de todos los valores considerados.

Se considera una variable de tamaño n que toma los valores x_1, x_2, \dots, x_n . La **media aritmética** de estos datos se define como

$$\tilde{x} = \frac{1}{n} \sum_{\ell=1}^n x_{\ell} \quad (13.1)$$

donde

- n : Representa el número total de observaciones. Por ejemplo, si se conoce el peso de 10 manzanas individuales, $n = 10$.
- X : La variable X es aquella sobre la que se calcula la media aritmética. En este caso sería el peso de las manzanas.
- ℓ : indexa cada observación. En este ejemplo, se podría asignar una etiqueta a cada manzana, la manzana 1, la manzana 2, etc.

De manera “informal”, a veces al resultado obtenido mediante la ecuación (13.1) se le llama **promedio**. Es el resultado de sumar el valor de la variable para todos los individuos y dividir por el total de individuos.

Para calcular la media aritmética de un conjunto de datos, es necesario disponer de los siguientes valores:

| | | | | |
|---|-------------------|-------------------|---------|-----------------------|
| Intervalos de clase a_j | $a_1 = e_0 - e_1$ | $a_2 = e_1 - e_2$ | \dots | $a_k = e_{k-1} - e_k$ |
| Frecuencia absoluta n_j | n_1 | n_2 | \dots | n_k |
| Frecuencia relativa f_j | f_1 | f_2 | \dots | f_k |

donde a_1, a_2, \dots, a_k son k intervalos, de manera que el intervalo a_j contiene n_j observaciones, siendo $n = \sum_{\ell=1}^k n_\ell$. La frecuencia absoluta de la clase j -ésima es $f_j = \frac{n_j}{n}$, y $\sum_{\ell=1}^n f_\ell = 1$.

El valor **medio** de la clase j se define como $m_j = \frac{e_{j-1} + e_j}{2}$, que se corresponde con la media del extremo superior y el extremo inferior, y se denomina la marca de clase.

La **media aritmética ponderada** es una medida de tendencia central, cuya consideración resulta apropiada cuando en un conjunto de datos cada uno de ellos tiene una importancia relativa (o peso) respecto de los demás datos. Para calcular la media aritmética ponderada en datos agrupados se utilizan las marcas de clase m_j previamente calculadas:

$$\tilde{x} = \frac{1}{n} \sum_{\ell=1}^n n_\ell m_\ell = \sum_{\ell=1}^n f_\ell m_\ell \quad (13.2)$$

Ejemplo 13.6.

Se consideran las temperaturas medidas en una ciudad durante el mes de mayo de 2020.

22, 24, 21, 22, 25, 26, 25, 24, 23, 25, 25, 26, 27, 25, 26,
25, 26, 27, 27, 28, 29, 29, 29, 28, 30, 29, 30, 31, 30, 28, 29

La temperatura media del mes de mayo, calculada como la media aritmética de los valores obtenidos es

$$\tilde{x} = \frac{22 + 24 + 21 + 22 + \dots + 31 + 30 + 28 + 29}{31} = 26,48^\circ C$$

Continuando con el mismo ejemplo, se supone que las distintas temperaturas están resumidas en distintas categorías como sigue:

| | | | | | |
|---|-----------|-----------------------|-----------------------|----------------------|-----------|
| Intervalos de clase a_j | < 20 | $(20, 25]$ | $(25, 30]$ | $(30, 35]$ | > 35 |
| Frecuencia absoluta n_j | $n_1 = 0$ | $n_2 = 12$ | $n_3 = 18$ | $n_4 = 1$ | $n_5 = 0$ |
| Frecuencia relativa f_j | $f_1 = 0$ | $f_2 = \frac{12}{31}$ | $f_3 = \frac{18}{31}$ | $f_4 = \frac{1}{31}$ | $f_5 = 0$ |

Estos valores permiten calcular la media aritmética ponderada:

$$\tilde{x} = \sum_{\ell=1}^k f_\ell m_\ell = 0 + \frac{12}{31} * 22,5 + \frac{18}{31} * 27,5 + \frac{1}{31} * 32,5 + 0 = 25,7$$

Es interesante observar que los resultados de la media y la media ponderada difieren. Esto es porque se ha usado el valor medio de cada clase como una aproximación de la media dentro de la clase. Esto sería correcto ante una asunción en general de distribución uniforme de los valores dentro de los intervalos, suposición que no se cumple. Si se conociera la media en cada clase, \tilde{x}_ℓ , como en este ejemplo, el resultado correcto se obtendría como:

$$\tilde{x} = \sum_{\ell=1}^k f_\ell \tilde{x}_\ell = 0 + \frac{12}{31} * 23,8333 + \frac{18}{31} * 28 + \frac{1}{31} * 32,5 + 0 = 26,48387$$

Sin embargo, la media ponderada está destinada a actuar como una estimación de la media aritmética en aquellos situaciones en las que sólo se dispone de datos agrupados. De hecho, en general se utiliza para obtener una aproximación de la media “real”. La diferencia entre una y otra se conoce como el error debido al agrupamiento

Propiedades de la media aritmética

1. La media aritmética sólo se puede calcular para variables numéricas o variables cualitativas dicotómicas codificadas como 0 y 1. En este caso, la media representa la frecuencia relativa de la categoría codificada como 1.
2. Un conjunto de datos numéricos sólo tiene una media.
3. La media es un parámetro sensible a la presencia de valores muy separados del resto de datos. Por ejemplo, la serie de valores, 1, 1, 2, 3, 3, 5, 7, 8, 8, 50 posee un valor extremo que es el 50. La media aritmética calculada con los 9 primeros valores es 4,2, lo que constituye un valor central razonable. Por el contrario, si se considera también el último valor, la media aritmética resulta ser 8,8, que es un valor muy poco indicativo del conjunto, pues está muy influido por ese valor extremo.
4. La suma de las desviaciones de cada variable alrededor de la media aritmética es cero

$$\sum_{\ell=1}^n (x_\ell - \tilde{x}) = \sum_{\ell=1}^n x_\ell - n\tilde{x} = n\tilde{x} - n\tilde{x} = 0. \quad (13.3)$$

Por tanto, la media es el centro de gravedad de la distribución de la variable.

5. Si los datos se transforman linealmente como $y_j = a + bx_i$, siendo a y b dos constantes conocidas, entonces se cumple:

$$\tilde{y} = \frac{1}{n} \sum_{\ell=1}^n y_\ell = \frac{1}{n} \sum_{\ell=1}^n (a + bx_\ell) = \frac{1}{n} \sum_{\ell=1}^n a + \frac{b}{n} \sum_{\ell=1}^n x_\ell = a + b\tilde{x} \quad (13.4)$$

Por tanto, los cambios de escala y origen en la variable cambian proporcionalmente y trasladan, respectivamente, a la media.

6. La media aritmética de los cuadrados de las desviaciones de los valores de la variable con respecto a una constante cualquiera se hace mínima cuando dicha constante coincide con la media aritmética.
7. La media aritmética está comprendida entre el valor máximo y el valor mínimo del conjunto de datos:

$$\min\{x_1, x_2, \dots, x_n\} \leq \frac{x_1 + x_2 + \dots + x_n}{n} \leq \max\{x_1, x_2, \dots, x_n\}$$

8. La media no tiene por qué ser igual a uno de los valores de los datos, ni siquiera de su misma naturaleza: datos enteros pueden tener una media decimal.
9. La media es una representación de los datos a partir de los que ha sido calculada, es decir, es un número que distingue un grupo de datos de otros (aunque es importante tener en cuenta medidas de dispersión para diferenciar grupos de datos con la misma media). En otros términos existe al menos un dato que es mayor o igual que la media aritmética.

Ejemplo 13.7.

Se retoma el ejemplo anterior acerca de las temperaturas en el mes de mayo. Estos datos se han medido en grados Celsius. Sin embargo, si quisieran mostrarse en la escala Fahrenheit, se podría llevar a cabo una transformación lineal para crear una nueva variable que represente la temperatura como:

$$F^o = 32 + 1,8 * ^o C$$

Por tanto,

$$\tilde{y} = a + b\tilde{x} = 32 + 1,8 * 26,48 = 79,7^o F$$

es el promedio de la temperatura de mayo en grados Fahrenheit.

Ejemplo 13.8.

En este ejemplo se plantea el cálculo de la media aritmética en una variable discreta. Se quiere calcular la media de edad de un conjunto de alumnos. Se añade a la tabla de frecuencias absolutas la columna con el producto de cada valor de la variable por su frecuencia.

| Valores | Frecuencia absoluta | $x_i f_i$ |
|--------------|---------------------|-------------|
| 12 | 9 | 108 |
| 13 | 25 | 325 |
| 14 | 27 | 378 |
| 15 | 16 | 240 |
| 16 | 12 | 192 |
| 17 | 8 | 136 |
| 18 | 3 | 54 |
| TOTAL | N = 100 | 1433 |

$$\tilde{x} = \frac{\sum_i x_i f_i}{\sum_i f_i} = \frac{1433}{100} = 14,33$$

La media de edad de los alumnos del centro entrevistados es de 14,33 años.

Aplicaciones de la media aritmética

Una de las aplicaciones importantes de la media aritmética es la de combinarse con medidas como la mediana y la moda para hacer comparaciones y analizar algunas distribuciones. La media aritmética tiene innumerables usos diferentes, muchos de ellos relacionados con aspectos de la vida cotidiana, como por ejemplo obtener promedio de una población a través de varios años; en las tablas financieras de interés compuesto y cuando se quiere dar importancia a valores pequeños. La progresión en el sistema educativo y la evaluación se basan en la media, el acceso a las titulaciones universitarias, los repartos económicos de los organismos internacionales, la consideración de los niveles de pobreza de los países, la presencia de obesidad o no en las personas, la rentabilidad de los activos financieros, etc.

Cuando tenemos que resumir un conjunto de datos numéricos es muy frecuente utilizar la media aritmética. La media aritmética o promedio destaca por representar el reparto equitativo, el equilibrio, la equidad. Es el valor que tendrían los datos, si todos ellos fueran iguales. O, también, el valor que correspondería a cada uno de los datos de la distribución si su suma total se repartiera por igual.

13.3.2 Media geométrica

Otra de las medidas de posición que se considera es la media geométrica. La media geométrica se calcula como un producto conjunto, habitualmente el de todos los valores de la media. De modo que si uno de ellos fuera cero, el producto total sería cero. Por ello, se debe tener en cuenta siempre que, a la hora de calcular la media geométrica, se deben considerar números que sean únicamente positivos.

Se consideran n observaciones de una variable cuantitativa, x_1, x_2, \dots, x_n , todos los valores

positivos. La **media geométrica** de estos datos, \bar{x}_G , se calcula como

$$\bar{x}_G = \sqrt[n]{\prod_{i=1}^n x_i} = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}} = \sqrt[n]{x_1 * x_2 * \dots * x_n} \quad (13.5)$$

donde

- n : Es el número total de observaciones. Por ejemplo, si se conoce el crecimiento de los beneficios de una empresa durante 4 periodos, $n = 4$.
- X : La variable X es aquella sobre la que se calcula la media geométrica. Siguiendo el ejemplo anterior, el crecimiento de los beneficios estará expresado en porcentaje y será la variable X .
- i : indexa cada observación. En este ejemplo, se le puede asignar un número cada periodo. Un 1, al periodo 1, un 2 al periodo 2, etc. De manera que x_1 es el crecimiento de los beneficios en el periodo 1, x_2 el crecimiento de los beneficios en el periodo 2, x_3 el crecimiento de los beneficios en el periodo 3 y x_4 el crecimiento de los beneficios en el periodo 4.

Ejemplo 13.13.9.

Se conocen los siguientes resultados de una empresa. La empresa ha generado un 20% de rentabilidad el primer año, un 15% el segundo año, un 33% el tercer año y un 25% el cuarto año. Lo fácil, en este caso sería sumar las cantidades y dividir entre cuatro. Sin embargo esto no es correcto. Para calcular la media de varios porcentajes debemos hacer uso de la media geométrica.

$$\tilde{x}_G = \sqrt[4]{1,21,151,331,25} = 1,23$$

El resultado es 1,23, que expresado en porcentaje es un 23%. Lo que quiere decir que en promedio, cada año la empresa ha ganado un 23%. Dicho de otra forma, si cada año hubiese ganado un 23%, hubiera ganado lo mismo que ganando un 20% el primer año, un 15% el segundo, un 33% el tercero y un 25% el último año.

Si las rentabilidades fueran negativas, no se pondrían números negativos en el cálculo de la media geométrica. En su lugar, se incorporarían valores < 1 que representen la misma situación. Por ejemplo, si la rentabilidad es del -20%, el número a considerar sería 0,80. Si la rentabilidad es del -5%, el número a multiplicar sería 0,95. En conclusión si las rentabilidades son positivas, a uno le sumamos el porcentaje en tanto por uno. Mientras que, si las rentabilidades o porcentajes son negativos, a 1 le restamos el porcentaje en tanto por uno.

Además del uso de la media geométrica para formalizar porcentajes, esta media juega un papel importante en los análisis relativos al producto de observaciones, como por ejemplo, al observar cambios porcentuales en cantidades. Es recomendada para datos de progresión geométrica, para promediar razones, interés compuesto y números índice. Su interpretación y uso se pueden entender como el crecimiento promedio de una cantidad, en el sentido de que existe un valor

inicial, como una cierta cantidad de dinero o una población particular, que cambian con el tiempo sobre este valor inicial. En el instante t , este valor podría haber cambiado, y por tanto pasaría a ser denotado como B_t , para $t = 1, 2, \dots, T$. La relación de B_t y B_{t-1} se llama t -ésimo **factor de crecimiento**, y se calcula como

$$x_t = \frac{B_t}{B_{t-1}}.$$

Por otra parte, el **ratio de crecimiento** se calcula como

$$r_t = ((x_t - 1) * 100) \%$$

El ratio de crecimiento proporciona una idea acerca del crecimiento o decrecimiento del valor de interés en el instante t . Todos estos conceptos se resumen en la Tabla 13.1.

| Instante t | Inventario B_t | Factor de crecimiento x_t | Ratio de crecimiento r_t |
|-----------------|---------------------|--------------------------------|-------------------------------|
| 0 | B_0 | - | - |
| 1 | B_1 | $x_1 = \frac{B_1}{B_0}$ | $((x_1 - 1) * 100) \%$ |
| 2 | B_2 | $x_2 = \frac{B_2}{B_1}$ | $((x_2 - 1) * 100) \%$ |
| ... | ... | ... | ... |
| T | B_T | $x_T = \frac{B_T}{B_{T-1}}$ | $((x_T - 1) * 100) \%$ |

Tabla 13.1: Elementos media geométrica

Para cada instante $t = 1, 2, \dots, T$, B_t se puede calcular utilizando el factor de crecimiento:

$$B_t = B_0 * x_1 * x_2 * \dots * x_t$$

El factor de crecimiento promedio de B_0 a B_T es la media geométrica, o media geométrica de los factores de crecimiento:

$$\bar{x}_G = \sqrt[n]{x_1 * x_2 * \dots * x_T} = \sqrt[n]{\frac{B_0 * x_1 * x_2 * \dots * x_T}{B_0}} = \sqrt[n]{\frac{B_T}{B_0}} \quad (13.6)$$

Por tanto, en el instante t , B_t se puede calcular como $B_t = B_0 * \bar{x}_G^t$

Ejemplo 13.10.

Se supone que alguien quiere depositar dinero en un banco, por ejemplo, 1000 €. El asesor bancario propone un plan de ahorro a 5 años con el siguiente plan de intereses: 1 % en el primer año; 1,5 % en el segundo año; 2,5 % en el tercer año, y 3 % en los últimos 2 años. Al cliente le interesaría calcular el factor de crecimiento promedio y la tasa de crecimiento promedio del dinero invertido. El concepto de media geométrica puede ser utilizado de la siguiente manera:

| Año | € | Factor de crecimiento | Tasa de crecimiento (%) |
|-----|---------|-----------------------|-------------------------|
| 0 | 1000 | — | — |
| 1 | 1010 | 1.01 | 1.0 |
| 2 | 1025.15 | 1.015 | 1.5 |
| 3 | 1050.78 | 1.025 | 2.5 |
| 4 | 1082.30 | 1.03 | 3.0 |
| 5 | 1114.77 | 1.03 | 3.0 |

En este caso, la media geométrica se calcula como

$$\bar{x}_G = (1,01 * 1,015 * 1,025 * 1,03 * 1,03)^{\frac{1}{5}} = 1,021968$$

Este valor de la media geométrica significa que habrá un crecimiento promedio de alrededor del 2,2% por año. Así, los ahorros después de 5 años se pueden calcular como

$$1000 * 1,021968^{\frac{1}{5}} = 1114,77€$$

Se puede observar que la media geométrica permite comparar fácilmente dos planes de ahorro con diferentes estrategias de crecimiento.

Propiedades de la media geométrica

1. El logaritmo de la media geométrica es igual a la media aritmética de los logaritmos de los valores de la variable.
2. La media geométrica de un conjunto de números positivos es siempre menor o igual que la media aritmética:

$$(x_1 x_2 \dots x_n)^{\frac{1}{n}} \leq \frac{x_1 + x_2 + \dots + x_n}{n}$$

La igualdad sólo se alcanza si $x_1 = x_2 = \dots = x_n$.

Una característica de la media geométrica es que considera todos los valores de la distribución. Además, es menos sensible que la media aritmética a los valores extremos. Sin embargo, tiene la desventaja de que, además de que su cálculo no es trivial, su significado estadístico e interpretación resultan menos intuitivos que los de la media aritmética. Por otra parte, es importante tener en cuenta que si hay un valor $x_i = 0$, entonces la media geométrica se anula o no queda determinada.

El cálculo de la media geométrica sólo es relevante si todos los números son positivos. Como se ha mencionado, si uno de ellos es 0, entonces el resultado es 0. Si hubiera un número negativo (o una cantidad impar de ellos) entonces la media geométrica sería o bien negativa, o bien inexistente en los números reales.

Al igual que con la media aritmética, la media geométrica también se puede calcular de manera **ponderada**, introduciendo pesos como valores multiplicativos para cada uno de los valores, con el fin de ponderar o hacer pesar más en el resultado final ciertos valores. En la media geométrica, los n pesos considerados, $\alpha_i, i = 1, 2, \dots, n$, se introducen como exponentes:

$$\bar{x}_G = \left(\prod_{\ell=1}^n x_{\ell}^{\alpha_{\ell}} \right)^{\frac{1}{\sum_{\ell=1}^n \alpha_{\ell}}} = (x_1^{\alpha_1} * x_2^{\alpha_2} * \dots * x_n^{\alpha_n})^{\frac{1}{\alpha_1 + \alpha_2 + \dots + \alpha_n}} \quad (13.7)$$

Aplicaciones de la media geométrica

Algunas de las aplicaciones más típicas de la media geométrica están relacionadas con la geometría en los siguientes casos:

- La altura de un triángulo rectángulo cumple $h = \sqrt{mn}$, siendo m y n las proyecciones de los catetos sobre la hipotenusa. Entonces, la altura es la media geométrica de los tamaños de las proyecciones de los catetos sobre la hipotenusa.
- Un cateto b cumple $b = \sqrt{ma}$, siendo m su proyección y a la hipotenusa del triángulo rectángulo.
- La tangente t a una circunferencia $t = \sqrt{sk}$, s es secante y k la parte interna.
- El lado de un cuadrado equivalente a un rectángulo es la media geométrica de los lados de este.
- El radio de un círculo equivalente a una elipse es la media geométrica de los semiejes de ésta. Lo mismo el caso de la esfera con el elipsoide.
- El lado (arista) d de un cubo equivalente a un ortoedro de lados a, b, c es $d = \sqrt[3]{abc^2}$

La media geométrica también aparece con asiduidad en problemas relativos a balanzas y pesas: en problemas relativos a crecimientos proporcionales, análisis en ciencias sociales, cálculos de relaciones de aspecto y coeficientes de tonalidad.

13.3.3 Media armónica

La media armónica se define como el *recíproco de la media aritmética de los recíprocos*.

La media armónica se usa típicamente siempre que hay diferentes valores x_i que contribuyan a la media cada uno de ellos con un peso diferente, w_i . Implícitamente, esto conduce a la asunción de que el peso asignado de cada x_i es distinto de uno. Así, dados k valores x_1, x_2, \dots, x_k , la media armónica puede calcularse como

$$\bar{x}_H = \frac{w_1 + w_2 + \dots + w_k}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \dots + \frac{w_k}{x_k}} = \frac{\sum_{\ell=1}^k w_{\ell}}{\sum_{\ell=1}^k \frac{w_{\ell}}{x_{\ell}}} = \frac{k}{\sum_{\ell=1}^k \frac{1}{x_{\ell}}} = \frac{k}{\frac{1}{x_1} + \dots + \frac{1}{x_k}}. \quad (13.8)$$

Por ejemplo, al calcular la velocidad media al recorrer una distancia n , cada peso se relaciona con la distancia recorrida en cada tramo n_i , a velocidad x_i . Usando como factor de peso $w_i = \frac{n_i}{n}$, y $\sum_{\ell} w_{\ell} = \sum_{\ell} \frac{n_{\ell}}{n}$, la media armónica puede calcularse como

$$\bar{x}_H = \frac{1}{\sum_{\ell=1}^k \frac{w_{\ell}}{x_{\ell}}} \quad (13.9)$$

Ejemplo 13.11.

Un inversor compró acciones por valor de 1000€ durante dos meses. El precio de una acción fue de 50€ en el primer mes, y de 200€ en el segundo. ¿Cuál es el precio medio de compra de una acción? El número de acciones compradas en el primer mes es $\frac{1000}{50} = 20$. El número de acciones compradas en el segundo mes es $\frac{1000}{200} = 5$. El número total de acciones compradas es, por tanto, $20 + 5 = 25$, y la inversión total es de 2000€. Es evidente que el precio medio de compra es $\frac{2000}{25} = 80$ €. De hecho, este es el valor de la media armónica calculada como

$$\bar{x}_H = \frac{1}{\frac{0,5}{50} + \frac{0,5}{200}} = 80$$

siendo el peso de cada compra $\frac{n_i}{n} = \frac{1000}{2000} = 0,5$. Si la inversión fue de 1200€ el primer mes, y 800€ en el segundo mes, se podría usar la media armónica con pesos $w_1 = \frac{1200}{2000} = 0,6$ y $w_2 = \frac{800}{2000} = 0,4$ para obtener el resultado.

Ejemplo 13.12.

Se supone que una familia realiza un viaje en automóvil a un ciudad y cubre los primeros 100 kilómetros a 60km/h , los siguientes 100 kilómetros a 70km/h y los últimos 100 kilómetros a 80km/h . Calcular, en esas condiciones, la velocidad media realizada.

$$\bar{x}_H = \frac{1}{\frac{1}{3} \left(\frac{1}{60} + \frac{1}{70} + \frac{1}{80} \right)} = 69,041 \text{ km/h}$$

Ejemplo 13.13.

Imaginemos que tenemos 4000€ para invertir en la acción ACME. No tenemos ni idea de qué hará la acción en el corto plazo, pero tenemos seguridad en ella de cara al largo plazo. Queremos evitar, eso sí, el ruido y el riesgo de comprar en máximos. Decidimos por lo tanto comprar una parte a final de cada trimestre del año.

Los precios de ACME a final de cada trimestre son:

- Trimestre 1: 8€.
- Trimestre 2: 4€.
- Trimestre 3: 5€.
- Trimestre 4: 10€.

Si se calcula la media aritmética, el precio obtenido sería

$$\tilde{x} = \frac{8 + 4 + 5 + 10}{4} = 7,75\text{€}$$

Si se compran 1000€ al final de cada trimestre, por un total de 4000€ se obtiene un total de 675 títulos como se puede observar:

- Trimestre 1: 1000€ / 8€ = 125 títulos
- Trimestre 2: 1000€ / 4€ = 250 títulos
- Trimestre 3: 1000€ / 5€ = 200 títulos
- Trimestre 4: 1000€ / 10€ = 100 títulos
- TOTAL = 125 + 250 + 200 + 100 = 675

Si se hubieran comprado a precio medio, tendríamos $\frac{4000\text{€}}{6,75\text{€}} = 592$ títulos, menos de los 675 que, efectivamente, están en cartera. Por lo tanto hemos comprado a un precio inferior al precio medio: 5,92€, valor que coincide con la media armónica.

$$\bar{x}_H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} = \frac{4}{\frac{1}{8} + \frac{1}{4} + \frac{1}{5} + \frac{1}{8}} = 5,92\text{€}$$

Ejemplo 13.14.

Se propone otro ejemplo relacionado con el cálculo del PER de una cartera, considerando las acciones ACME y EMCA:

- ACME. Precio=10€; beneficio por acción: 1€; por lo tanto, $\text{PER} = \frac{10}{1} = 10$.
- EMCA: Precio=8€; beneficio por acción: 2€; por lo tanto, $\text{PER} = \frac{8}{2} = 4$.
- PER medio:

– Media aritmética: $\tilde{x} = \frac{10+4}{2} = 4$

– Media armónica: $\bar{x}_H = \frac{\text{suma de precios de las acciones}}{\text{suma de beneficio por acción}} = \frac{10+8}{1+2} = 6$

Propiedades de la media armónica

1. La inversa de la media armónica es la media aritmética de los inversos de los valores de la variable.

2. Siempre se puede pasar de una media armónica a una media aritmética transformando adecuadamente los datos.
3. La media armónica siempre es menor o igual que la media aritmética, ya que para cualquier número real positivo $x_i > 0$ se cumple

$$\frac{k}{\frac{1}{x_1} + \dots + \frac{1}{x_k}} \leq \frac{x_1 + \dots + x_k}{k}$$

Una de las ventajas de la media armónica es que considera todos los valores de la distribución y en ciertos casos, es más representativa que la media aritmética. Además, puede representar con un valor muy pequeño el promedio de un conjunto de números con valores muy grandes.

Por el contrario, tiene la desventaja de la influencia de los valores pequeños y el hecho de que no pueda ser determinada en distribuciones con valores iguales a cero. Esto implica que su empleo no sea aconsejable en distribuciones donde existan valores muy pequeños, ya que sus inversos pueden aumentar mucho, haciendo despreciable frente a ellos la información de otros valores de la variable en estudio que sean mayores. Además, se requiere una capacidad de cálculo mayor para obtener la media armónica que para calcular otras medidas de centralidad. Es importante también tener en cuenta que no debe usarse para valores de una variable muy pequeños (cerca de 0).

Aplicaciones de la media armónica

El uso de la media armónica no está muy extendido en el mundo científico. Esta medida se utiliza principalmente para calcular la media de velocidades, tiempos o en electrónica.

También se utiliza en determinados problemas de física, de recuperación de información, hidrología, automovilística y geometría.

Bibliografía

1. Rustom Jabbaz, A. (2012). *Estadística descriptiva, probabilidad e inferencia. Una visión conceptual y aplicada*. Santiago de Chile: Departamento de Economía Agraria, Facultad de Ciencias Agronómicas, Universidad de Chile.
2. Heumann, C. y Schomaker Shalabh, M. (2016). *Introduction to Statistics and Data Analysis, with Exercises, Solutions and Applications in R*. Springer.

Tema 14

Estadística descriptiva III. Medidas de posición robustas. Mediana, moda y cuantiles. Cálculo de las mismas y propiedades. Aplicaciones.

Este tema está elaborado como una adaptación de la siguiente bibliografía:

Venancio Tomeo Perucha e Isaías Uña Juárez (2009). *Estadística descriptiva*. Madrid: Ibergarceta Publicaciones

Esta documentación es orientativa y no es exclusiva ni única para el correcto desarrollo de este tema. Tampoco vincula al órgano convocante ni al Tribunal actuante.

Aviso: El INE se reserva las acciones legales que pudieran corresponder por la venta de esta información.

14.1 Introducción

En este capítulo vamos a estudiar las medidas de posición robusta. A diferencia de la media aritmética, que es la primera medida de posición que se enseña en los cursos de estadística, las medidas de posición robustas ofrecen la ventaja de no ser sensibles a los valores extremos de la variable.

Recordemos que la media aritmética es una medida de posición central que nos sirve para representar todos los datos. Sin embargo, la media es sensible a la presencia de valores extremos. Por ejemplo, supongamos que analizamos los datos de la cotización de las empresas durante el tercer trimestre del año 2021 y entre las empresas que se analizan: algunas se dedican a la venta de mascarillas (quirúrgicas, FFP2, FFP3...); otras son tecnológicas (ofrecen servicios online); otras del sector restauración; del sector de la automoción; etc. Si no tenemos en cuenta las distintas agrupaciones por sector económico, es probable que tengamos valores extremos en nuestros datos. En este caso, la media no sería representativa y sería preferible utilizar otras alternativas, como son las medidas de posición robustas. Podemos emplear la mediana, si deseamos continuar con una medida de posición central, o la moda y cuantiles, si deseamos utilizar otras medidas.

La Figura 14.1 presenta la distribución del efectivo en el Mercado Continuo en un día concreto.

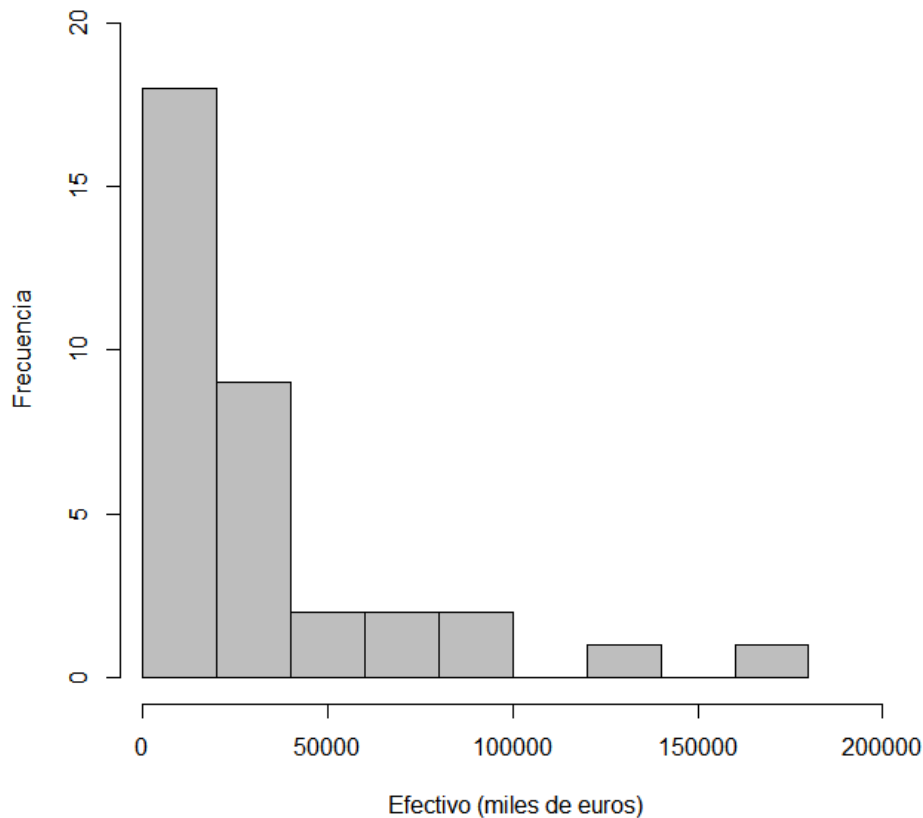


Figura 14.1: Histograma del Efectivo (Mercado continuo)

A la vista del histograma, una pregunta importante a responder es si la media puede ser representativa de los datos. En caso negativo, ¿qué alternativas tenemos?

Las medidas de posición robustas se calculan para variables cuantitativas. Se pueden aplicar en diversas áreas: economía, psicología, medicina...

14.2 Mediana

La mediana de una distribución es aquel número tal que, ordenados los datos de forma creciente (de menor a mayor), la mitad son inferiores a dicho valor (las observaciones situadas a la izquierda) y la otra mitad son superiores (las observaciones situadas a su derecha). La mediana se denota por M_e .

Podemos considerar dos tipos de distribuciones cuando los datos no están agrupados: una distribución de frecuencias unitarias y una distribución de frecuencias no unitarias. El cálculo de la mediana es diferente en cada caso.

14.2.1 Distribuciones de frecuencias unitarias: número de observaciones impares

En las distribuciones sin agrupar y N impar, existe un único valor de la variable en el centro de la distribución, y este valor coincide con la mediana.

Ejemplo 1. Las temperaturas más bajas registradas durante el mes de julio en Zaragoza, en grados centígrados, son las siguientes:

| Temperatura de Zaragoza |
|-------------------------|
| 22 |
| 12 |
| 23 |
| 30 |
| 25 |

Si ordenamos los valores de la distribución de menor a mayor, el diagrama en escalera (o acumulativo) quedaría de la siguiente forma:

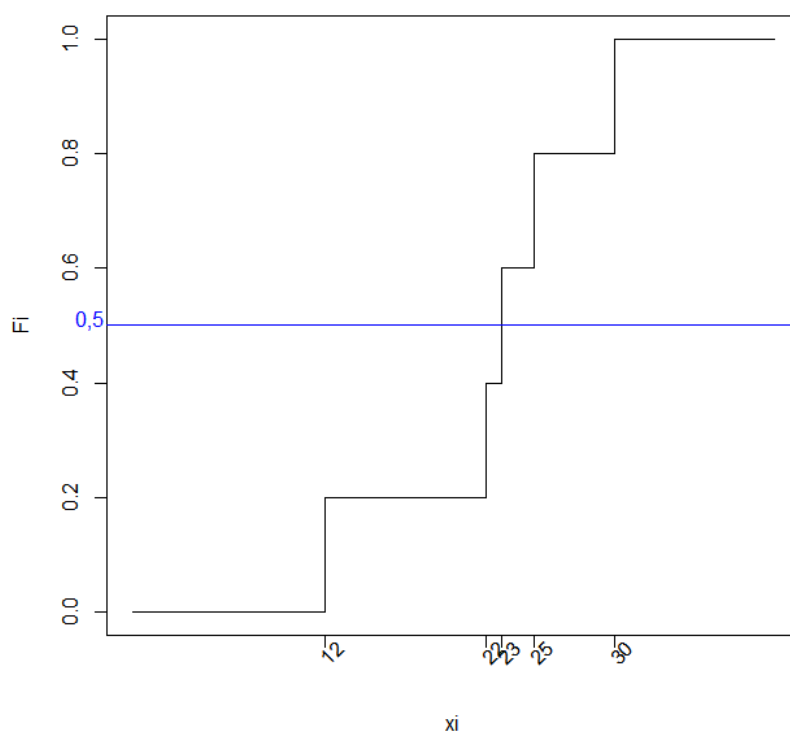


Figura 14.2: Diagrama en escalera de la temperatura en Zaragoza

A la vista del diagrama anterior, observamos que el valor que deja a su izquierda al menos el 50 % de las observaciones ordenadas es el 23. Es decir, la mediana es 23 grados centígrados.

14.2.2 Distribuciones de frecuencias unitarias: número de observaciones pares

En las distribuciones sin agrupar y N par, la mediana se define como la media aritmética de los dos valores centrales de la distribución:

$$M_e = \frac{x_i + x_{i+1}}{2}.$$

Este es el criterio que se va a emplear para calcular la mediana. Pero es posible considerar otros criterios como $M_e = x_i$, $M_e = x_{i+1}$, u otro valor comprendido entre estos valores anteriores.

Ejemplo 2. Los kilómetros recorridos por 6 corredores durante su entrenamiento para un maratón se presentan en la siguiente tabla:

| Kilómetros recorridos |
|-----------------------|
| 500 |
| 730 |
| 120 |
| 375 |
| 600 |
| 1000 |

Si ordenamos los valores de la distribución de menor a mayor, el diagrama escalonado presenta la siguiente forma:

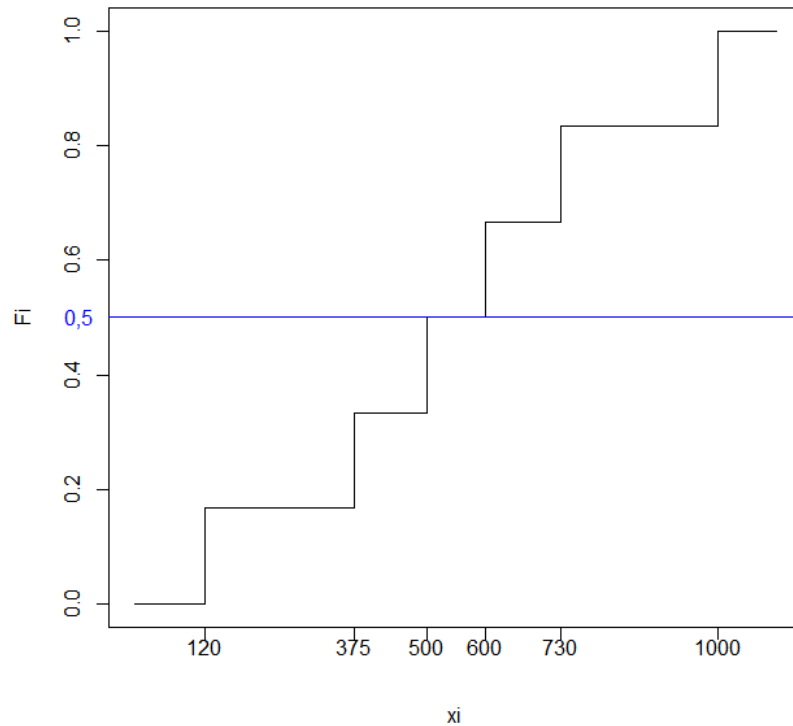


Figura 14.3: Diagrama en escalera de los kilómetros recorridos

La mediana viene representada por

$$M_e = \frac{500 + 600}{2} = 550 \text{ km.}$$

14.2.3 Distribuciones de frecuencias no unitarias

Cuando la distribución de frecuencias no es unitaria, primero, se construye la tabla de frecuencias en la que aparezcan las frecuencias absolutas acumuladas (o relativas acumuladas); segundo, se observa el primer valor de la variable para el cual la frecuencia absoluta acumulada sea igual o superior a $N/2$ (o la frecuencia relativa igual o superior a 0,5); y tercero, se calcula la mediana como en el caso de las distribuciones de frecuencias unitarias, dependiendo de si N es impar o par.

Si N_i es la primera frecuencia absoluta acumulada igual o superior a $N/2$, entonces

$$\begin{cases} M_e = x_i, & \text{si } N_{i-1} < \frac{N}{2} < N_i, \\ M_e = \frac{x_i + x_{i+1}}{2}, & \text{si } N_i = \frac{N}{2}. \end{cases}$$

Ejemplo 3. Los metros recorridos por 20 lanzas en una competición deportiva se presentan en la siguiente tabla:

| Metros recorridos | n_i | N_i |
|-------------------|-------|-------|
| 100 | 7 | 7 |
| 102 | 5 | 12 |
| 105 | 8 | 20 |

La mediana es 102 metros. Ya que si nos fijamos en las frecuencias absolutas acumuladas, $7 < \frac{N}{2} = 10 \leq 12$.

Gráficamente, podemos observar que la mediana es 102 metros.

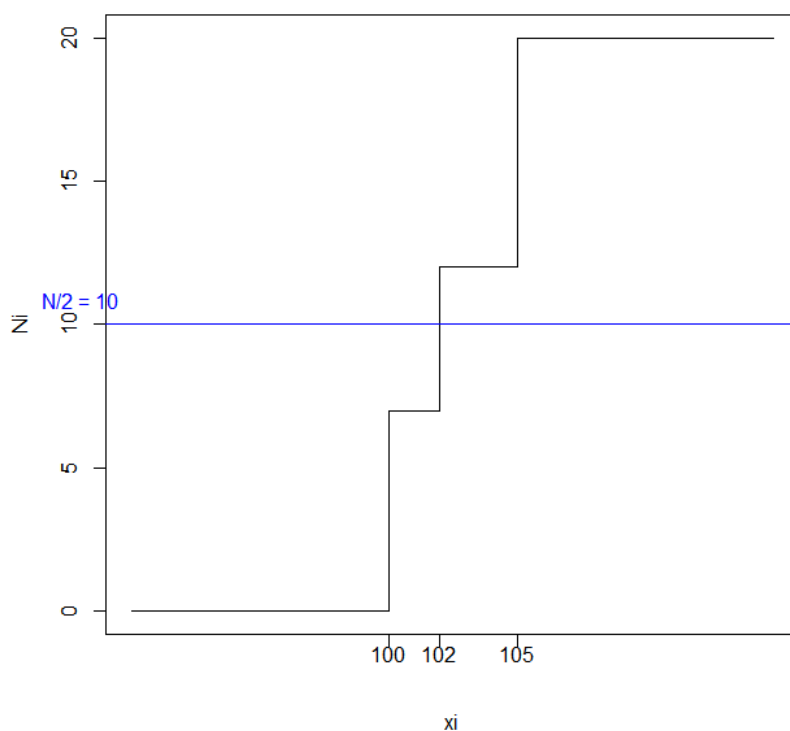


Figura 14.4: Diagrama en escalera de los metros recorridos

Ejemplo 4. Ahora supongamos que los metros recorridos por 24 lanzas en una competición deportiva se presentan en la siguiente tabla:

| Metros recorridos | n_i | N_i |
|-------------------|-------|-------|
| 500 | 7 | 7 |
| 505 | 5 | 12 |
| 511 | 12 | 24 |

La mediana es 508 metros. Si nos fijamos en las frecuencias absolutas acumuladas, $\frac{N}{2} = 12$ y coincide con la frecuencia absoluta acumulada de N_2 . Por tanto, la mediana es

$$M_e = \frac{505 + 511}{2} = 508 \text{ metros.}$$

Gráficamente, podemos observar que la mediana es 102 metros.

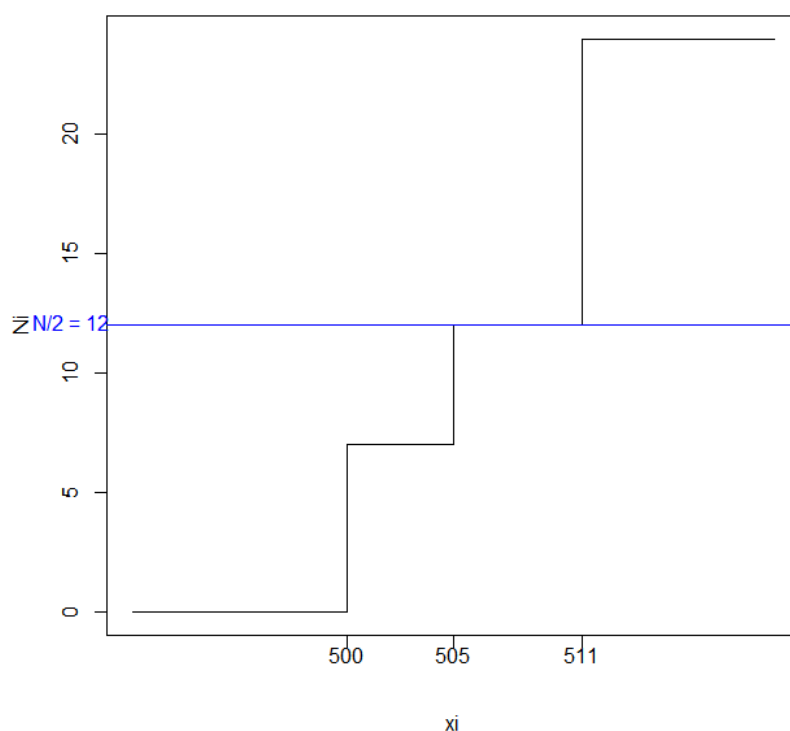


Figura 14.5: Diagrama en escalera de los metros recorridos

14.2.4 Distribución de frecuencias agrupadas

En el caso de distribuciones de frecuencias agrupadas, el problema se resuelve buscando el intervalo mediano. El intervalo mediano es el primer intervalo donde la frecuencia

acumulada alcanza o sobrepasa $N/2$. Para precisar el valor de mediana en este tipo de distribuciones, debemos suponer que la frecuencia correspondiente al intervalo se distribuye uniformemente en la amplitud del intervalo y repartiendo proporcionalmente se obtiene el valor de la mediana:

$$M_e = L_{i-1} + c_i \frac{\frac{N}{2} - N_{i-1}}{n_i},$$

donde:

L_{i-1} es el extremo inferior del intervalo mediano,

N_{i-1} es la frecuencia absoluta acumulada en el intervalo anterior al mediano,

n_i es la frecuencia absoluta del intervalo mediano (también se puede calcular como $N_i - N_{i-1}$),

c_i es la amplitud del intervalo mediano,

N es el número total de datos.

Ejemplo 5. Una socorrista tiene que salvar a 24 muñecos en el océano Atlántico. El tiempo en minutos que ha tardado en rescatar a los muñecos ha sido agrupado de la siguiente forma:

| Tiempo | Muñecos rescatados |
|----------|--------------------|
| [1, 7) | 4 |
| [7, 15) | 7 |
| [15, 25] | 5 |
| [25, 60] | 8 |

Primero vamos a construir la columna de las frecuencias absolutas acumuladas:

| Tiempo (L_{i-1}, L_i) | Muñecos rescatados (n_i) | N_i |
|---------------------------|------------------------------|-------|
| [1, 7) | 4 | 4 |
| [7, 15) | 7 | 11 |
| [15, 25) | 5 | 16 |
| [25, 60] | 8 | 24 |

También podemos representar los datos de la tabla anterior mediante el polígono acumulativo de frecuencias.

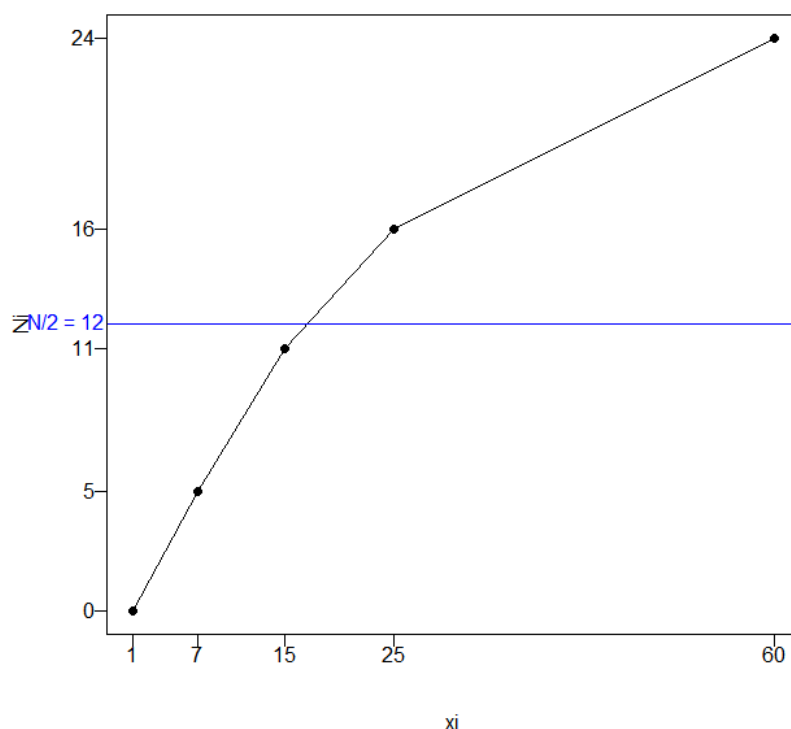


Figura 14.6: Polígono acumulativo de frecuencias del tiempo de rescate

El intervalo mediano es $[15, 25)$, puesto que $11 < \frac{N}{2} = 12 \leq 16$. Una vez identificado el intervalo mediano podemos obtener el valor de la mediana del siguiente modo:

| Tiempo (L_{i-1}, L_i) | Muñecos rescatados (n_i) | N_i | Amplitud (c_i) |
|---------------------------|------------------------------|-------|--------------------|
| [1, 7) | 4 | 4 | 6 |
| [7, 15) | 7 | 11 | 8 |
| [15, 25) | 5 | 16 | 10 |
| [25, 60] | 8 | 24 | 35 |

$$M_e = 15 + 10 \frac{\frac{24}{2} - 11}{5} = 17 \text{ minutos.}$$

14.3 Moda

La moda de una distribución es el valor de la variable con mayor frecuencia. La moda se denota por M_o . Para determinar la moda, se procede como se hizo con la mediana: se distingue entre distribuciones de valores sin agrupar y agrupados.

14.3.1 Distribuciones de frecuencias de valores sin agrupar

En las distribuciones de frecuencias de valores sin agrupar se calcula la moda fijándose en el valor de la variable que más se repite (es decir, el de mayor frecuencia). Para ello, en la tabla de frecuencias nos fijamos en la columna de frecuencias absolutas. En este tipo de distribuciones, la moda no tiene por qué ser única. Las distribuciones con más de un valor modal se conocen como distribución multimodal (hay otras denominaciones específicas: bimodal, dos modas; trimodal, tres modas; etc.).

En las siguientes figuras se representan una distribución unimodal y otra bimodal:

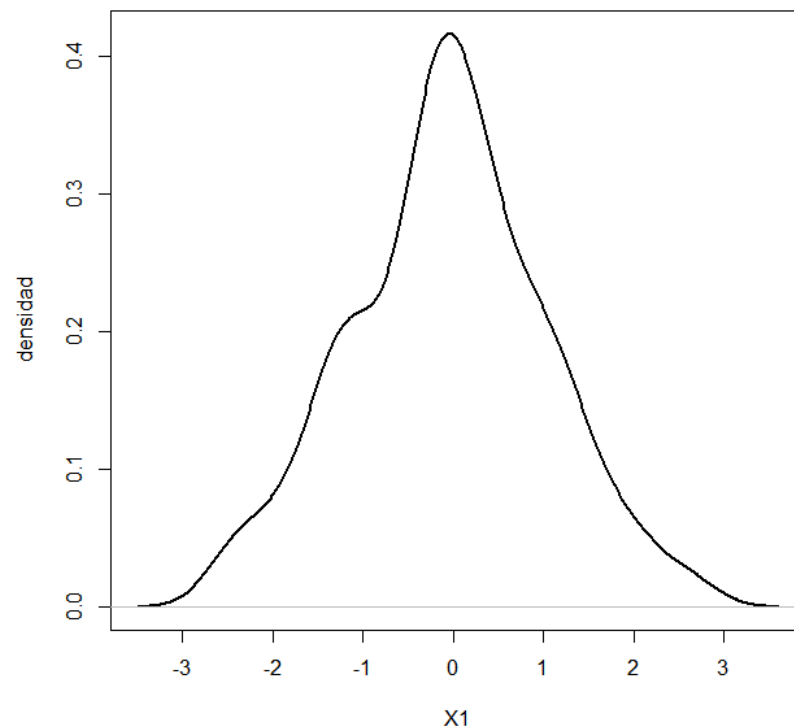


Figura 14.7: Distribución unimodal

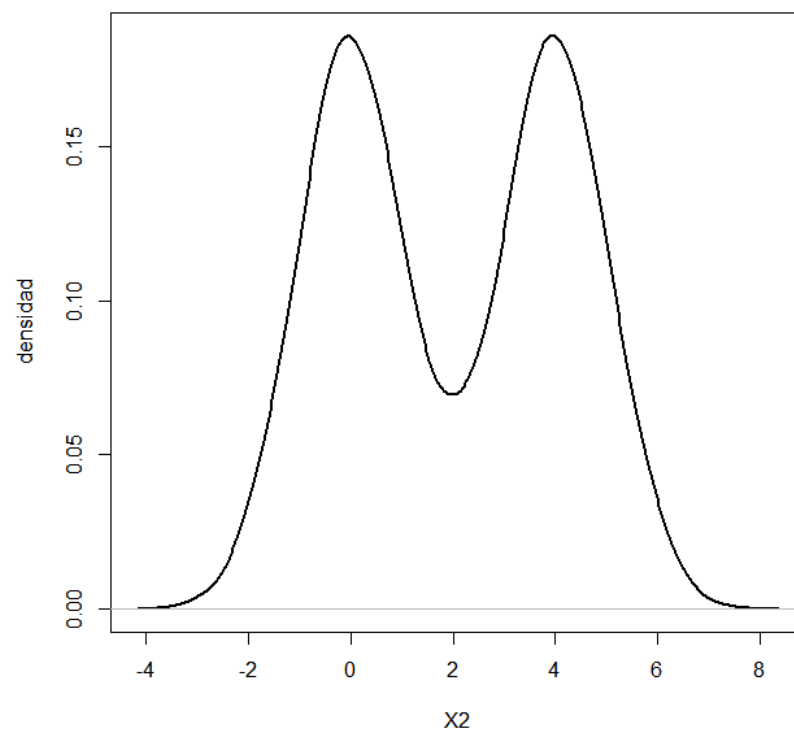


Figura 14.8: Distribución bimodal

Ejemplo 6. Retomando los datos del Ejemplo 3, podemos representar el diagrama de barras de los datos:

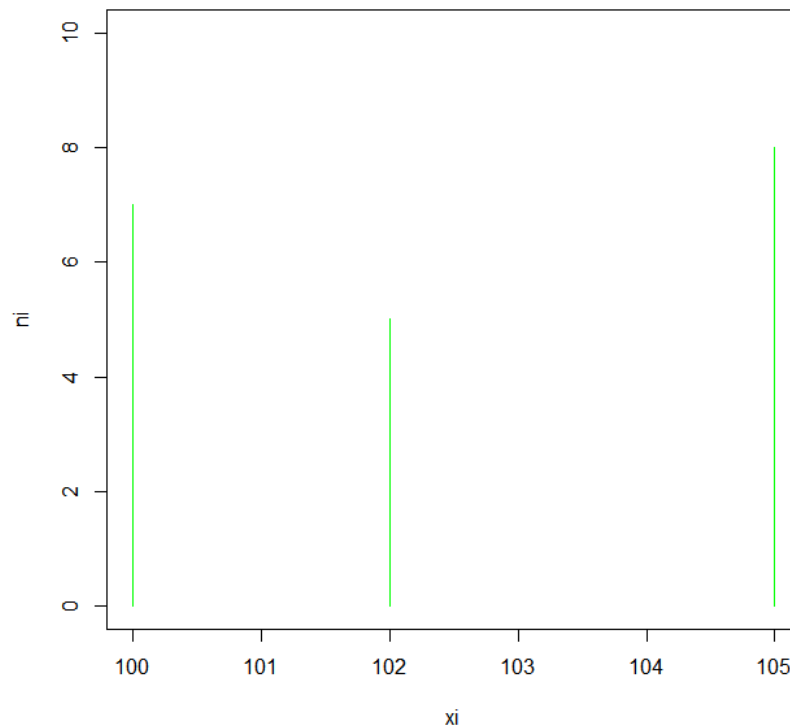


Figura 14.9: Diagrama de barras de los metros recorridos

A la vista de las frecuencias absolutas, podemos comprobar que los metros recorridos más frecuentes son 105 metros. Es decir, la moda de la variable es 105 metros (máximo n_i o f_i).

14.3.2 Distribuciones de frecuencias de valores agrupados

En el caso de distribuciones de frecuencias agrupadas, el problema se resuelve buscando el intervalo modal. La moda se encontrará en el intervalo de mayor altura, que se denomina intervalo modal. Para determinar la altura se calcula el cociente entre la frecuencia absoluta y la amplitud. La moda se obtiene como

$$M_o = L_{i-1} + c_i \frac{h_{i+1}}{h_{i-1} + h_{i+1}},$$

donde:

L_{i-1} es el extremo inferior del intervalo modal,

h_{i-1} es la altura asociada al intervalo anterior al modal,

h_{i+1} es la altura asociada al intervalo posterior al modal,

c_i es la amplitud del intervalo mediano.

La altura, h_i , se calcula como el cociente entre n_i y c_i .

Ejemplo 7. Retomando los datos del Ejemplo 5, el intervalo modal es $[7, 15)$, puesto que la altura máximo se produce en ese intervalo. Una vez identificado el intervalo modal podemos obtener el valor de la moda del siguiente modo:

| Tiempo (L_{i-1}, L_i) | n_i | N_i | Amplitud (c_i) | Altura (h_i) |
|-------------------------|-------|-------|------------------|----------------|
| $[1, 7)$ | 4 | 4 | 6 | 0,667 |
| $[7, 15)$ | 7 | 11 | 8 | 0,875 |
| $[15, 25)$ | 5 | 16 | 10 | 0,500 |
| $[25, 60]$ | 8 | 24 | 35 | 0,229 |

$$M_o = 7 + 8 \frac{0,500}{0,667 + 0,500} = 10,428,$$

Si representamos las alturas, observamos que el intervalo modal es $[7, 15)$.

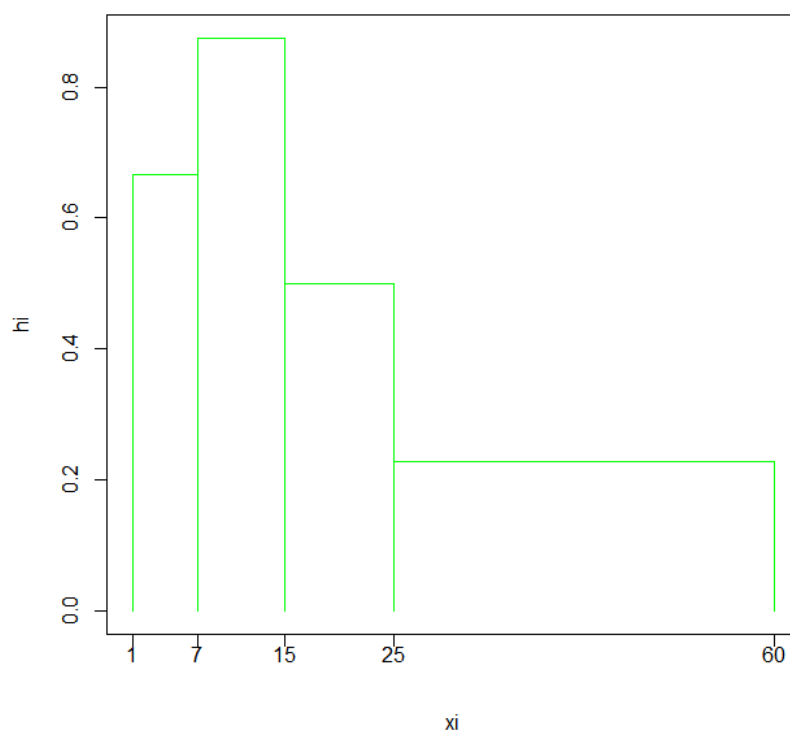


Figura 14.10: Altura h_i del tiempo de rescate

En el caso de que las alturas coincidan, la anterior y posterior a la modal, la moda es el valor central del intervalo modal (es decir, la marca de la clase del intervalo modal). Si la altura del intervalo anterior al modal es cero, la moda es L_i . Si la altura del intervalo

posterior al modal es cero, la moda es L_{i-1} . Cuando los intervalos de la distribución tienen la misma amplitud, podemos sustituir en el cálculo de la moda la altura por la frecuencia absoluta:

$$M_o = L_{i-1} + c_i \frac{n_{i+1}}{n_{i-1} + n_{i+1}},$$

donde:

L_{i-1} es el extremo inferior del intervalo modal,

n_{i-1} es la frecuencia absoluta asociada al intervalo anterior al modal,

n_{i+1} es la frecuencia absoluta asociada al intervalo posterior al modal,

c_i es la amplitud del intervalo mediano.

Ejemplo 8. Hállese la moda de la siguiente distribución estadística:

| L_{i-1}, L_i | n_i |
|----------------|-------|
| [1, 5) | 10 |
| [5, 9) | 20 |
| [9, 13) | 27 |
| [13, 17) | 5 |
| [17, 21] | 15 |

Representamos el histograma de nuestros datos:

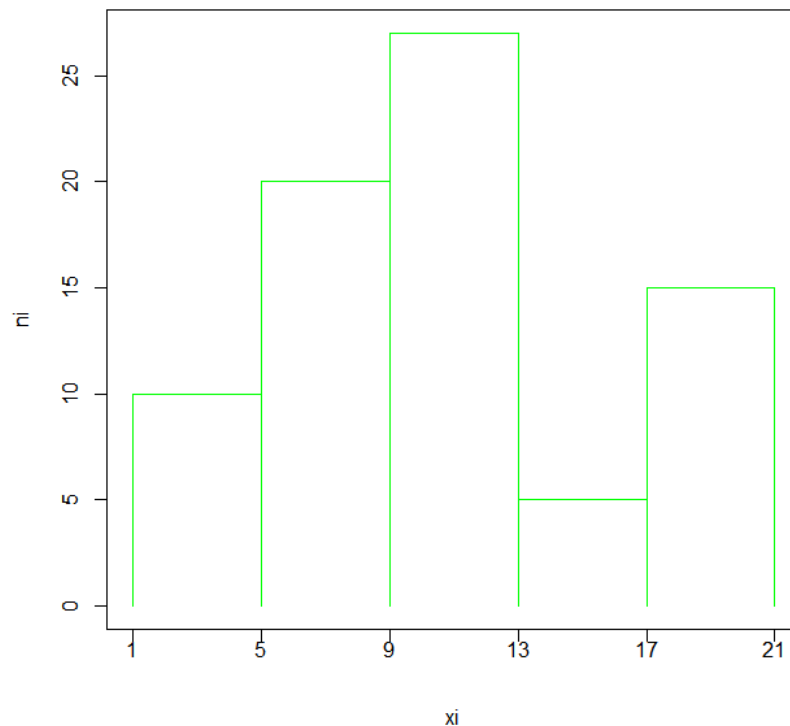


Figura 14.11: Histograma de los datos

El intervalo modal es $[9, 13)$, puesto que la frecuencia absoluta máxima se produce en ese intervalo. Una vez identificado el intervalo modal podemos obtener el valor de la moda como

$$M_o = 9 + 4 \frac{5}{5 + 20} = 9,8.$$

El valor de la moda será inferior, igual o superior a la marca de clase del intervalo modal, dependiendo de si las alturas de los intervalos anterior y posterior al modal. Sean I_1 el intervalo anterior al modal e I_2 el intervalo posterior al modal, entonces

- (a) Si la frecuencia absoluta (o altura) de I_1 es mayor que la frecuencia absoluta (o altura) de I_2 , $M_o < \text{Marca de clase del intervalo modal}$.
- (b) Si la frecuencia absoluta (o altura) de I_1 es igual a la frecuencia absoluta (o altura) de I_2 , $M_o = \text{Marca de clase del intervalo modal}$.
- (c) Si la frecuencia absoluta (o altura) de I_1 es menor que la frecuencia absoluta (o altura) de I_2 , $M_o > \text{Marca de clase del intervalo modal}$.

Si en nuestro ejemplo cambiamos las frecuencias absolutas de los intervalos anterior y posterior al modal del siguiente modo

| L_{i-1}, L_i | n_i |
|----------------|-------|
| [1, 5) | 10 |
| [5, 9) | 5 |
| [9, 13) | 27 |
| [13, 17) | 20 |
| [17, 21] | 15 |

Y representamos el histograma de los datos:

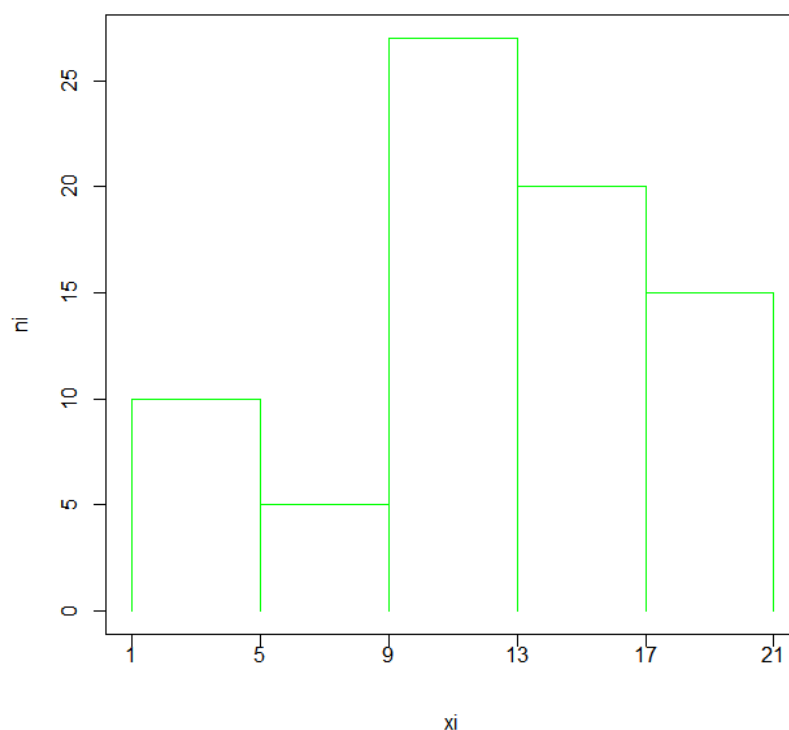


Figura 14.12: Histograma de los datos (*cambios en n_i)

El intervalo modal sigue siendo $[9, 13)$, puesto que la frecuencia absoluta máxima se produce en ese intervalo. Pero ahora el valor de la moda es:

$$M_o = 9 + 4 \frac{20}{20 + 5} = 12,2.$$

En este caso, la moda es superior a la marca de clase del intervalo modal (11).

Por último, un procedimiento menos utilizado es calcular la moda encontrando la parábola que pasa por los puntos medios de las bases superiores de los rectángulos del

histograma correspondiente al intervalo modal, al anterior y al posterior. La moda es la abscisa de su vértice.

14.4 Cuantiles

De forma análoga a la mediana, los cuantiles son aquellos valores de la distribución de la variable que dividen a la distribución en k partes, conteniendo cada una de ellas la misma proporción de observaciones ($\frac{1}{k}$), habiendo sido previamente ordenados de menor a mayor los valores de la distribución. En general, si dividimos la distribución en k partes, siendo k un valor entero positivo, $C_{\frac{s}{k}}$ es el cuantil de orden ($\frac{s}{k}$), de tal forma que, al menos, el $(\frac{s}{k} * 100)\%$ de los valores de la variable serán menores o iguales a este cuantil y, al menos, el $(1 - \frac{s}{k}) * 100\%$ de los valores de la variable serán iguales o mayores a este cuantil. En total, se obtienen $(k - 1)$ valores de la variable que contendrán una proporción de valores de ($\frac{1}{k}$). Por tanto, $s = 1, 2, \dots, (k - 1)$.

En el cálculo de los cuantiles, se procede de forma análoga a la mediana, distinguiendo entre datos sin agrupar (con N par o impar) y datos agrupados. Para datos sin agrupar:

$$\begin{cases} C_{\frac{s}{k}} = x_i, & \text{si } N_{i-1} < \frac{s \cdot N}{k} < N_i, \\ C_{\frac{s}{k}} = \frac{x_i + x_{i+1}}{2}, & \text{si } N_i = \frac{s \cdot N}{k}, \end{cases}$$

donde $C_{\frac{s}{k}}$ es el cuantil que se desea calcular y que contendrá una proporción de ($\frac{s}{k}$) observaciones.

En el caso de datos agrupados, se procede a encontrar el primer intervalo donde acumula al menos el $(\frac{s}{k})\%$ observaciones. Para calcular el valor del cuantil, podemos hacer la misma suposición que en el caso de la mediana, donde hemos supuesto que la frecuencia correspondiente al intervalo se distribuye uniformemente en la amplitud del intervalo y repartiendo proporcionalmente se obtiene el valor del cuantil:

$$C_{\frac{s}{k}} = L_{i-1} + c_i \frac{\frac{s \cdot N}{k} - N_{i-1}}{n_i},$$

donde $s = 1, 2, \dots, k - 1$.

La familia de cuantiles más utilizadas son los cuartiles, los deciles y los percentiles.

14.4.1 Cuantiles

Los cuantiles son tres valores que separan la distribución en cuatro partes iguales, conteniendo cada una de ellas el 25 % de las observaciones. El primer cuartil es valor de la distribución tal que, al menos, el 25 % de los datos son iguales o inferiores a él y el 75 % restantes son, al menos, iguales o superiores. Se denota por Q_1 . El segundo cuartil es valor de la distribución tal que, al menos, el 50 % de los datos son iguales o inferiores a él. Se denota por Q_2 . Como puede verse este valor coincide con la mediana. El tercer cuartil es valor de la distribución tal que, al menos, el 75 % de los datos son iguales o inferiores a él y el 25 % restantes son, al menos, iguales o superiores. Se denota por Q_3 . En el cálculo de los cuantiles, se procede de forma análoga a la mediana, distinguiendo entre datos sin agrupar (con N par o impar) y datos agrupados. En este caso, $k = 4$ y $s = 1, 2, 3$.

Ejemplo 9. Los kilómetros recorridos por 15 corredores durante su entrenamiento para un maratón se presentan en la siguiente tabla:

| Kilómetros recorridos | n_i | N_i |
|-----------------------|-------|-------|
| 120 | 2 | 2 |
| 375 | 5 | 7 |
| 500 | 3 | 10 |
| 600 | 1 | 11 |
| 730 | 3 | 14 |
| 1000 | 1 | 15 |

El diagrama escalonado presenta la siguiente forma:

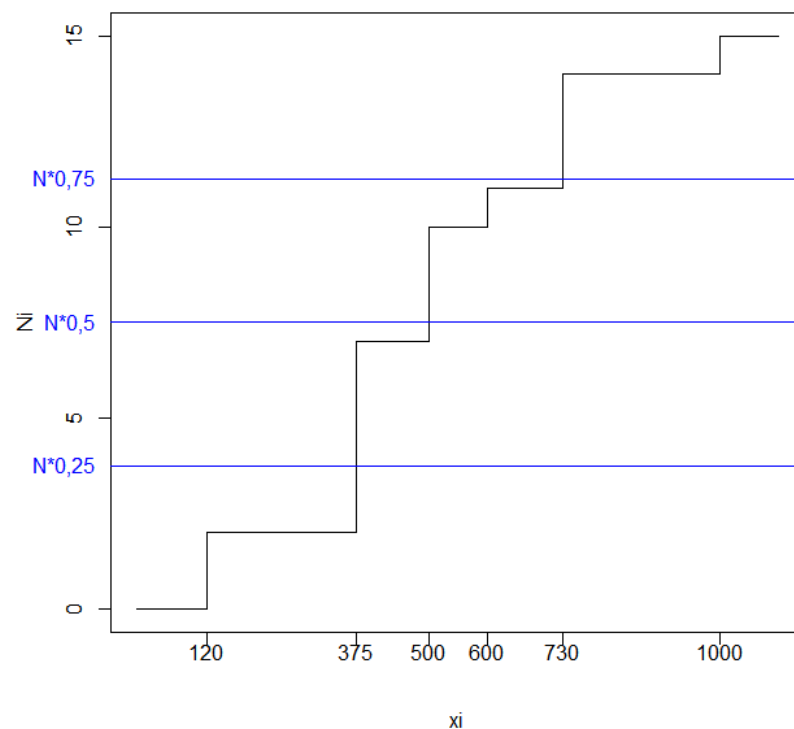


Figura 14.13: Diagrama en escalera de los kilómetros recorridos

El primer cuartil es 375 km, el segundo cuartil es 500 y el tercer cuartil es 730. Por ejemplo, el primer cuartil nos indica que el 25 % de los corredores hacen un recorrido igual o inferior a 375 km.

Ejemplo 10. Retomando los datos del Ejemplo 8, podemos calcular los tres cuantiles de la distribución:

| L_{i-1}, L_i | n_i | N_i |
|----------------|-------|-------|
| [1, 5) | 10 | 10 |
| [5, 9) | 20 | 30 |
| [9, 13) | 27 | 57 |
| [13, 17) | 5 | 62 |
| [17, 21] | 15 | 77 |

También podemos representar los datos de la tabla anterior mediante el polígono acumulativo de frecuencias.

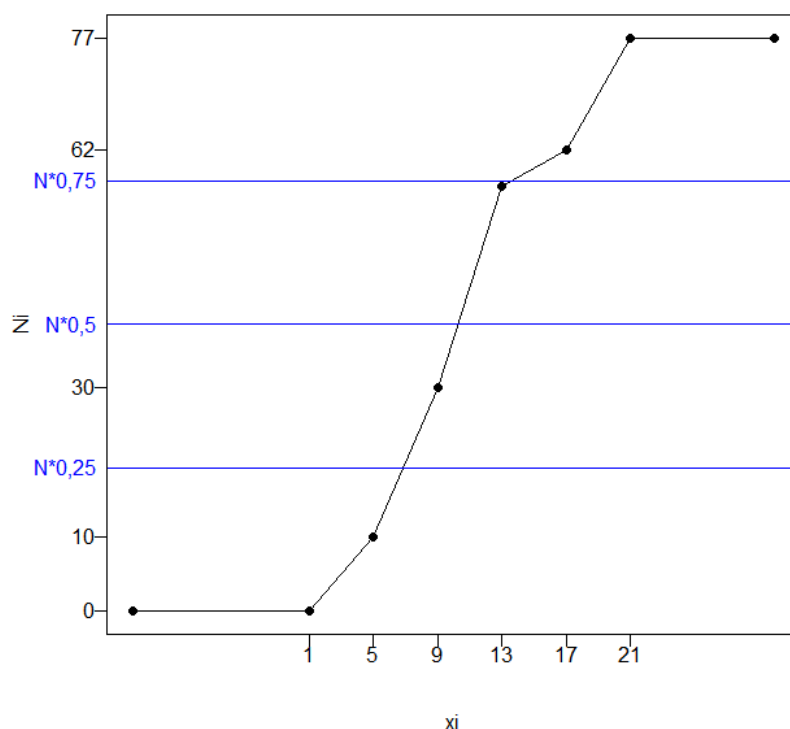


Figura 14.14: Polígono acumulativo de frecuencias de los datos

El intervalo del primer cuartil es $[5, 9)$, del segundo cuartil es $[9, 13)$ y del tercer cuartil es $[13, 17)$.

Si queremos saber cuál es el valor del primer cuartil, Q_1 , podemos obtenerlo como

$$Q_1 = 5 + 4 \frac{\frac{1 \cdot 77}{4} - 10}{20} = 6,85.$$

14.4.2 Deciles

Los deciles son nueve valores que separan la distribución en diez partes iguales, conteniendo cada una de ellas el 10 % de las observaciones. Se denotan por D_1, D_2, \dots, D_9 . Por ejemplo, el primer decil es valor de la distribución tal que, al menos, el 10 % de los datos son iguales o inferiores a él y el 90 % restantes son, al menos, iguales o superiores. En el cálculo de los deciles, se procede de forma análoga a la mediana, distinguiendo entre datos sin agrupar (con N par o impar) y datos agrupados. En este caso, $k = 10$ y $s = 1, 2, \dots, 9$.

Ejemplo 11. Retomando los datos del Ejemplo 9, vamos a calcular los deciles de la distribución:

| Kilómetros recorridos | n_i | N_i |
|-----------------------|-------|-------|
| 120 | 2 | 2 |
| 375 | 5 | 7 |
| 500 | 3 | 10 |
| 600 | 1 | 11 |
| 730 | 3 | 14 |
| 1000 | 1 | 15 |

El diagrama escalonado presenta la siguiente forma:

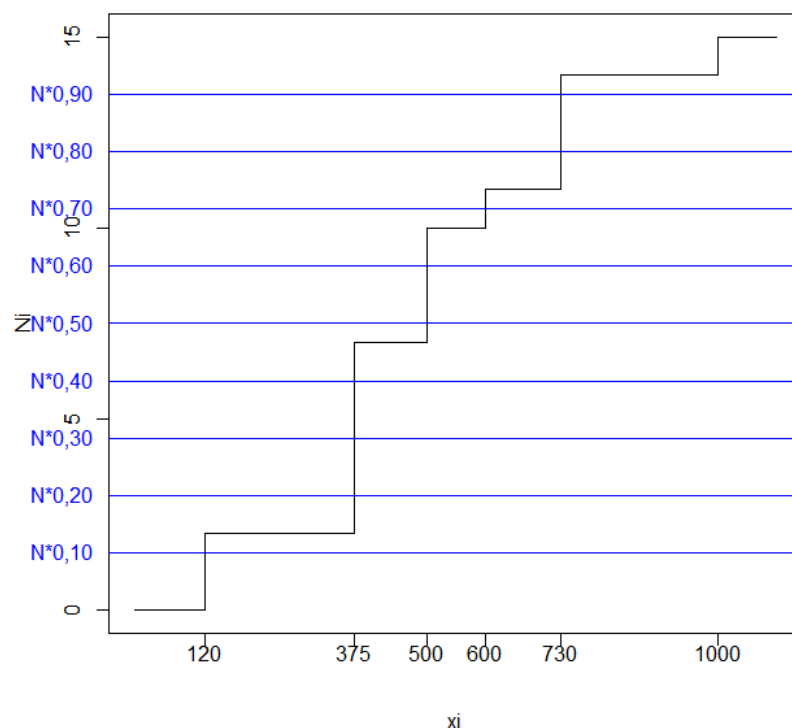


Figura 14.15: Diagrama en escalera de los kilómetros recorridos

El primer decil es 120 km; el segundo, tercer y cuarto decil es 375 km; el quinto y sexto decil es 500; el séptimo decil es 600 km; y el octavo y noveno decil es 730 km. Por ejemplo, el primer decil nos indica que el 10 % de los corredores hacen un recorrido igual o inferior a 120 km.

Ejemplo 12. Retomando los datos del Ejemplo 8, podemos calcular los nueve deciles de la distribución:

| L_{i-1}, L_i | n_i | N_i |
|----------------|-------|-------|
| [1, 5) | 10 | 10 |
| [5, 9) | 20 | 30 |
| [9, 13) | 27 | 57 |
| [13, 17) | 5 | 62 |
| [17, 21] | 15 | 77 |

También podemos representar los datos de la tabla anterior mediante el polígono acumulativo de frecuencias.

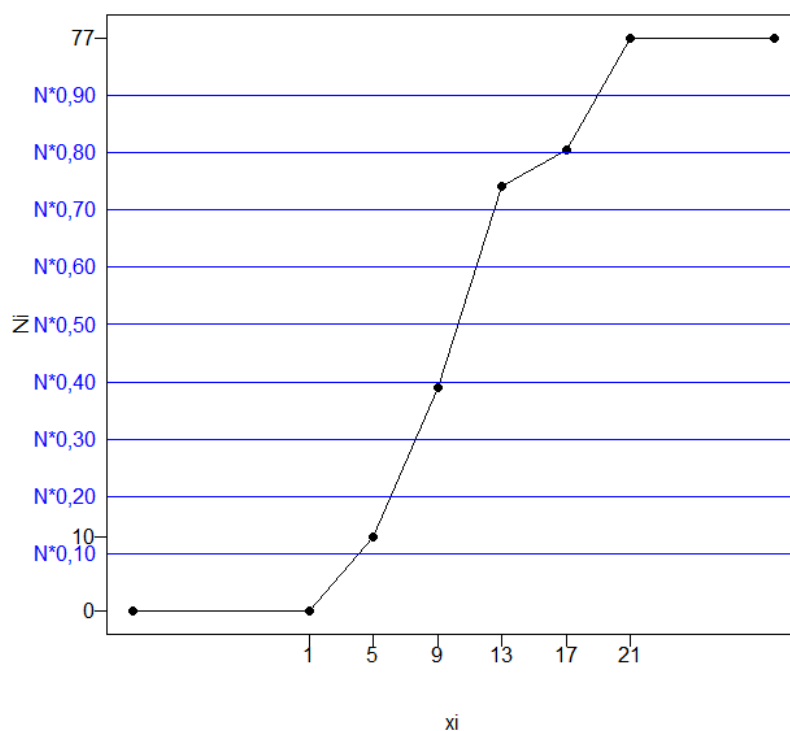


Figura 14.16: Polígono acumulativo de frecuencias de los datos

El intervalo del primer decil es [1, 5); del segundo y tercer decil es [5, 9); del cuarto, quinto, sexto y séptimo decil es [9, 13); del octavo decil es [13, 17); y del noveno decil es [17, 21].

Si queremos saber cuál es el valor del primer decil, D_1 , podemos obtenerlo como

$$D_1 = 1 + 4 \frac{\frac{1 \cdot 77}{10} - 0}{10} = 4,08.$$

14.4.3 Percentiles

Los percentiles son noventa y nueve valores que separan la distribución en cien partes iguales, conteniendo cada una de ellas el 1 % de las observaciones. Se denotan por P_1, P_2, \dots, P_{99} . Por ejemplo, el primer percentil es valor de la distribución tal que, al menos, el 1 % de los datos son iguales o inferiores a él y el 99 % restantes son, al menos, iguales o superiores. En el cálculo de los percentiles, se procede de forma análoga a la mediana, distinguiendo entre datos sin agrupar (con N par o impar) y datos agrupados. En este caso, $k = 100$ y $s = 1, 2, \dots, 99$.

Podemos establecer relaciones entre los cuartiles, deciles y percentiles, quedando de la siguiente forma:

$$D_1 = P_{10}$$

$$D_2 = P_{20}$$

$$Q_1 = P_{25}$$

$$D_3 = P_{30}$$

$$D_4 = P_{40}$$

$$Q_2 = D_5 = P_{50}$$

$$D_6 = P_{60}$$

$$D_7 = P_{70}$$

$$Q_3 = P_{75}$$

$$D_8 = P_{80}$$

$$D_9 = P_{90}$$

Ejemplo 13. Retomando los datos del Ejemplo 9, vamos a calcular algunos percentiles de la distribución:

| Kilómetros recorridos | n_i | N_i |
|-----------------------|-------|-------|
| 120 | 2 | 2 |
| 375 | 5 | 7 |
| 500 | 3 | 10 |
| 600 | 1 | 11 |
| 730 | 3 | 14 |
| 1000 | 1 | 15 |

El diagrama escalonado presenta la siguiente forma:

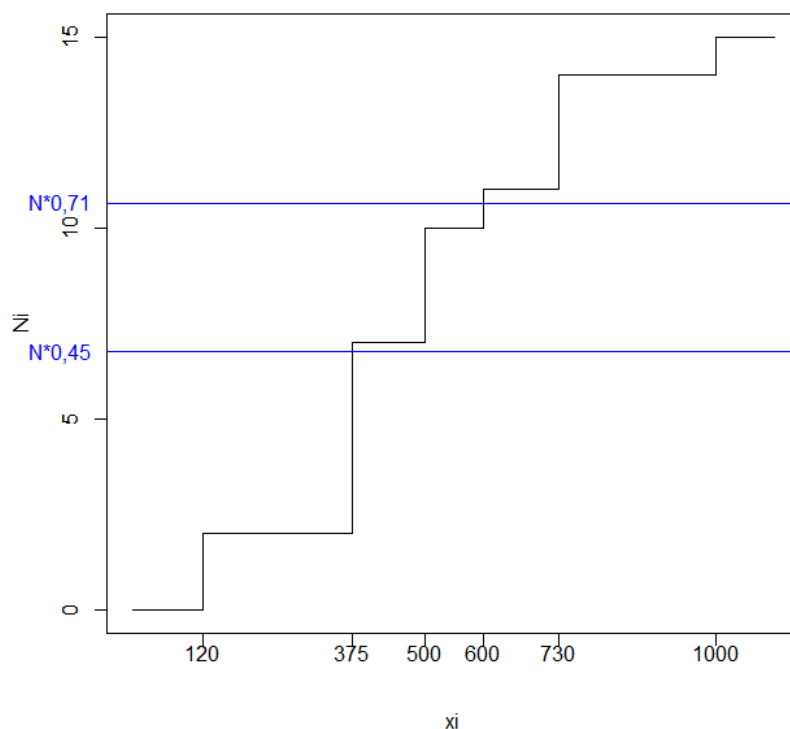


Figura 14.17: Diagrama en escalera de los kilómetros recorridos

$P_{45} = 375$ km. $P_{71} = 600$ km. Por ejemplo, el cuadragésimo quinto percentil nos indica que el 45 % de los corredores hacen un recorrido igual o inferior a 375 km.

Ejemplo 14. Retomando los datos del Ejemplo 8, vamos a calcular un percentil de la distribución:

| L_{i-1}, L_i | n_i | N_i |
|----------------|-------|-------|
| [1, 5) | 10 | 10 |
| [5, 9) | 20 | 30 |
| [9, 13) | 27 | 57 |
| [13, 17) | 5 | 62 |
| [17, 21] | 15 | 77 |

También podemos representar los datos de la tabla anterior mediante el polígono acumulativo de frecuencias.

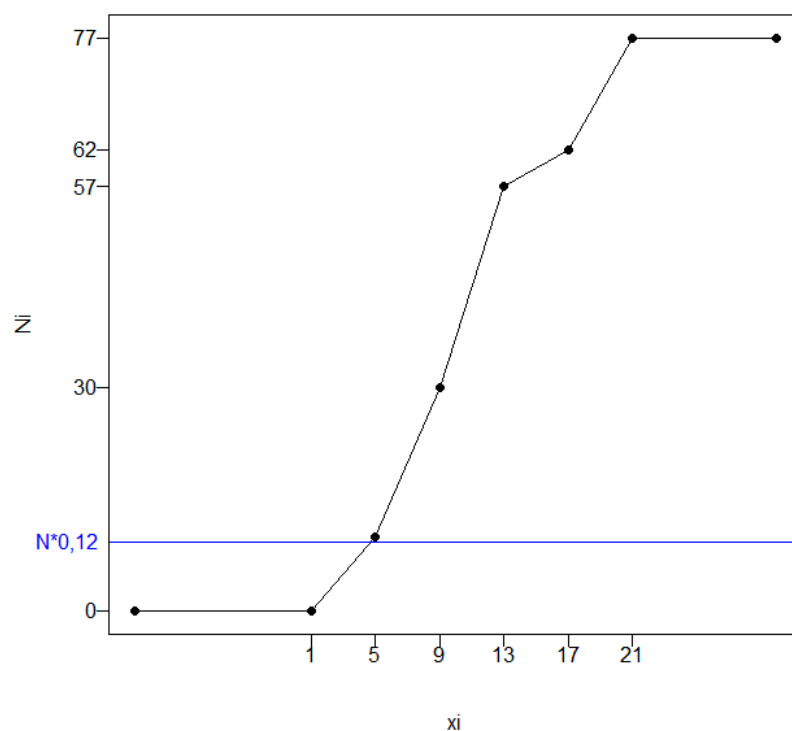


Figura 14.18: Polígono acumulativo de frecuencias de los datos

El intervalo del duodécimo percentil es $[1, 5)$. Si queremos saber cuál es el valor del duodécimo percentil, D_{12} , podemos obtenerlo como

$$P_{12} = 1 + 4 \frac{\frac{12 \cdot 77}{100} - 0}{10} = 4,696.$$

Problema propuesto. Un economista desea analizar la cotización de un grupo de empresas españolas. Para ello, se ha fijado en el precio de la acción en euros (X) y el efectivo en miles de euros (Y) en una sesión del Mercado Continuo. Los datos se presentan en la siguiente tabla:

| X | n_i | Y | n_j |
|--------|-------|------------------|-------|
| 0,54 | 1 | [2800, 5000) | 3 |
| 30,7 | 6 | [5000, 10000) | 7 |
| 78,8 | 4 | [10000, 15000) | 6 |
| 153,5 | 8 | [15000, 25000) | 4 |
| 244,5 | 9 | [25000, 40000) | 7 |
| 559,4 | 5 | [40000, 70000) | 3 |
| 1300,2 | 2 | [70000, 100000) | 3 |
| | | [100000, 170000] | 2 |

Se pide:

- ¿A qué valor máximo se cotizan las empresas del primer decil?
- ¿A qué precio se suele comprar la acción?
- ¿Cuál es el efectivo más frecuente?
- Un inversor quiere invertir en una cartera de acciones si la mediana del precio de la acción no supera los 200. Ante una cartera como la de tus datos, ¿qué hará el inversor?
- Si el inversor prefiere los mercados donde el efectivo mínimo del 20 % de las empresas cotizadas supere los 70 millones de euros, ¿qué hará ante un mercado como el propuesto?
- No se disponen de los datos del volumen, pero se sabe que es el cociente entre el efectivo y el precio de la acción. Haz una aproximación del volumen mediano a partir de las medianas calculadas en los apartados anteriores.

Por último, se recomienda la lectura de otros libros de varios autores para ampliar los conocimientos respecto de la materia estudiada ([Bachero Nebot y col. 2006](#); [Peña 2001](#)).

Bibliografía

- Peña, Daniel (2001). *Fundamentos de estadística*. Madrid: Alianza editorial.
- Bachero Nebot, José, Olga Blasco Blasco, Vicente Coll Serrano, Rafael Díez García, Jesús Esteban García, Antonia Ivars Escortell, María Isabel López Rodríguez, Concepción Rojo Olivas y Félix Ruiz Ponce (2006). *Estadística descriptiva y nociones de probabilidad*. Editorial Paraninfo.
- Tomeo Perucha, Venancio e Isaías Uña Juárez (2009). *Estadística descriptiva*. Madrid: Ibergarceta Publicaciones.

Tema 15

Estadística descriptiva IV. Medidas de dispersión. Recorrido, varianza y desviación típica. Otras medidas de dispersión. Cálculo de las mismas y propiedades. Aplicaciones.

Este tema está elaborado como una adaptación de la siguiente bibliografía:

tomeo2009estadistica

Esta documentación es orientativa y no es exclusiva ni única para el correcto desarrollo de este tema. Tampoco vincula al órgano convocante ni al Tribunal actuante.

Aviso: El INE se reserva las acciones legales que pudieran corresponder por la venta de esta información.

15.1 Introducción

La dispersión o variabilidad es una medida estadística que mide la proximidad o lejanía entre los datos de una distribución estadística. En este sentido, si los valores de la variable están próximos entre sí, diremos que los datos presentan poca dispersión; y si están alejados, diremos que hay mucha dispersión.

Las medidas de dispersión se puede agrupar en dos grandes grupos: medidas de dispersión absoluta y medidas de dispersión relativa.

Las medidas de dispersión absoluta más utilizadas son:

- Rango o recorrido
- Varianza
- Desviación típica

Las medidas de dispersión relativa tienen la ventaja de que permiten comparar la dispersión o variabilidad entre varias distribuciones de frecuencias. Las más comunes son:

- Coeficiente de variación
- Tipificación de variables

Existen numerosas medidas de dispersión. Pero algunos son de uso infrecuente como, por ejemplo, la desviación media, la desviación mediana, la MEDA, la desviación cuartílica y la desviación percentílica 10-90.

Tomando de ejemplo la distribución del efectivo en el Mercado Continuo en un día concreto, una forma de aproximarnos al cálculo de la dispersión de una distribución es preguntarnos: ¿cuál es la distancia de cada valor de la variable a la media? ¿y a la mediana? ¿o cuál es la distancia entre el valor más alto y el más bajo?

La Figura 15.1 presenta la distribución del efectivo en el Mercado Continuo en un día concreto y resalta algunos estadísticos.

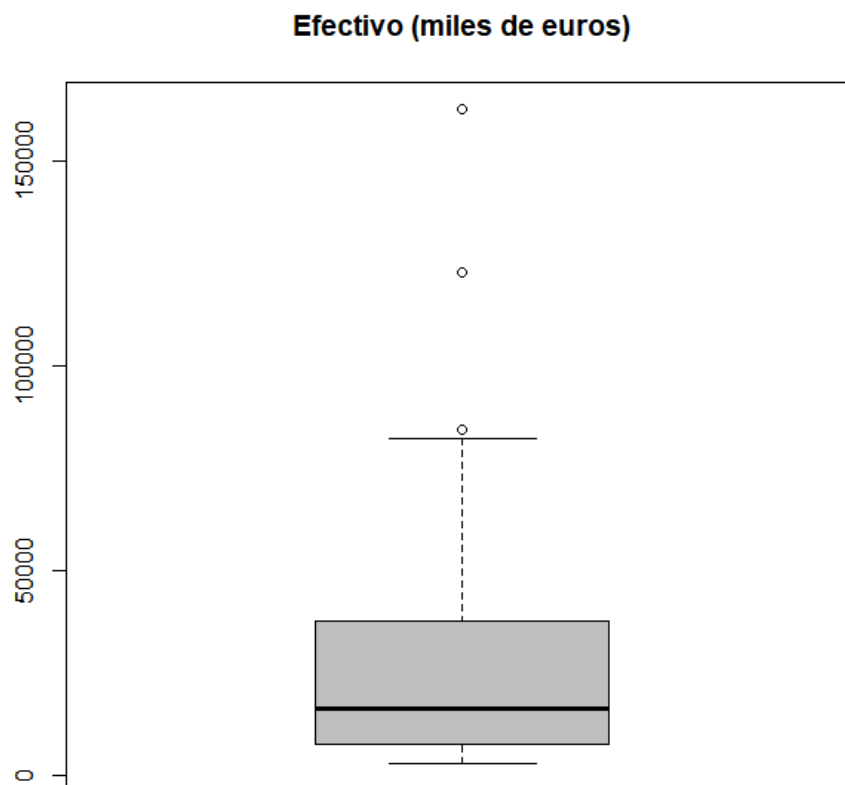


Figura 15.1: Diagrama de caja y bigotes del Efectivo (Mercado continuo)

En la tabla siguiente se presentan algunas medidas de posición de la variable.

| Estadístico | Valor del Efectivo (en miles de euros) |
|-------------|--|
| Mín. | 2858 |
| Q_1 | 7691 |
| Mediana | 16167 |
| Media | 31375 |
| Q_3 | 37542 |
| Máx. | 162571 |

A partir del diagrama de caja y de la tabla de estadísticos descriptivos, podemos responder a la cuestión de la variabilidad. Por ejemplo, fijándonos en la distancia entre el efectivo más alto y el más bajo.

Sin embargo, si quisiéramos comparar la dispersión de nuestra distribución con otra tomada en un día distinto, ¿podemos aprovechar la información de la tabla?. ¿Qué alternativas deberíamos utilizar?

15.2 Recorrido

El recorrido o rango de una distribución es la diferencia entre los valores máximo y mínimo de la variable. Se denota por R_e y viene expresado como

$$R_e = x_{max} - x_{min}.$$

Ejemplo 1. Las temperaturas más bajas registradas durante el mes de julio en Rabat en grados centígrados, se presentan en la siguiente tabla:

| Temperatura de Rabat |
|----------------------|
| 22 |
| 18 |
| 19 |
| 24 |
| 16 |

Para calcular el recorrido de la distribución primero ordenamos los valores de la variable de menor a mayor:

| Temperatura de Rabat (x_i ordenados) |
|---|
| 16 |
| 18 |
| 19 |
| 22 |
| 24 |

La temperatura mínima es de 16 grados centígrados y la máxima de 24 grados centígrados. Por tanto, el recorrido es 8 grados centígrados. Se obtiene como

$$R_e = 24 - 16 = 8 \text{ grados centígrados.}$$

15.3 Varianza

La varianza es una medida de dispersión que se define como la media aritmética de los cuadrados de las desviaciones de los datos respecto a su media aritmética. Se denota por S^2 , se obtiene como

$$S^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 * n_i}{N}.$$

Si en la fórmula anterior se desarrolla el cuadrado, podemos calcular la varianza como

$$S^2 = \frac{\sum_{i=1}^k x_i^2 * n_i}{N} - \bar{x}^2.$$

El valor de la varianza será siempre igual o superior a cero. Es decir, $S^2 \geq 0$.

Las expresiones anteriores se pueden simplificar cuando los valores de las variables sean todos distintos ya que sus respectivas frecuencias absolutas son iguales a 1 (es decir, $n_i = 1 \forall 1 \leq i \leq k$), y $k = N$. Las expresiones anteriores quedarían de la siguiente forma:

$$S^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N};$$

$$S^2 = \frac{\sum_{i=1}^N x_i^2}{N} - \bar{x}^2.$$

Ejemplo 2. Retomando los datos del Ejemplo 1, la varianza se puede calcular de las siguientes formas:

Alternativa 1.

| x_i |
|-------|
| 16 |
| 18 |
| 19 |
| 22 |
| 24 |

La media de la variable es 19,8 grados centígrados. Se obtiene como

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} = \frac{99}{5} = 19,8.$$

Y la varianza es

$$S^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N} = \frac{(22 - 19,8)^2 + (18 - 19,8)^2 + (19 - 19,8)^2 + (24 - 19,8)^2 + (16 - 19,8)^2}{5} = 8,16.$$

Alternativa 2.

Para obtener la varianza construimos la columna de los valores de la variable al cuadrado:

| x_i | x_i^2 |
|---------------|---------|
| 16 | 256 |
| 18 | 324 |
| 19 | 361 |
| 22 | 484 |
| 24 | 576 |
| $\sum = 2001$ | |

$$S^2 = \frac{\sum_{i=1}^N x_i^2}{N} - \bar{x}^2 = \frac{2001}{5} - 19,8^2 = 8,16.$$

También puede obtenerse directamente como

$$S^2 = \frac{22^2 + 18^2 + 19^2 + 24^2 + 16^2}{5} - 19,8^2 = \frac{2001}{5} - 19,8^2 = 8,16.$$

Por tanto, la varianza de la temperatura en Rabat es de 8,16 grados centígrados al cuadrado.

Cuando los datos estén agrupados, es necesario calcular previamente la marca de clase para hallar la varianza.

Ejemplo 3. Un enfermero tiene que vacunar a 24 personas en una hora. Las personas vacunadas en distintos tramos de tiempo se presentan en la siguiente tabla:

| Tiempo | Personas vacunadas |
|----------|--------------------|
| [1, 7) | 4 |
| [7, 15) | 7 |
| [15, 25] | 5 |
| [25, 60] | 8 |

Primero vamos a calcular la marca de clase para calcular la media y la varianza:

| Tiempo (L_{i-1}, L_i) | Marca de clase (x_i) | n_i | $x_i * n_i$ |
|---------------------------|--------------------------|----------|--------------|
| [1, 7) | 4 | 4 | 16 |
| [7, 15) | 11 | 7 | 77 |
| [15, 25) | 20 | 5 | 100 |
| [25, 60] | 42,5 | 8 | 340 |
| | | $N = 24$ | $\sum = 533$ |

La media es:

$$\bar{x} = \frac{\sum_{i=1}^k x_i * n_i}{N} = \frac{533}{24} = 22,2083.$$

De forma directa:

$$\bar{x} = \frac{\sum_{i=1}^k x_i * n_i}{N} = \frac{4 * 4 + 11 * 7 + 20 * 5 + 42,5 * 8}{24} = \frac{533}{24} = 22,2083.$$

Alternativa 1.

Para obtener la varianza construimos la siguiente columna en la tabla:

| Tiempo (L_{i-1}, L_i) | x_i | n_i | $x_i * n_i$ | $(x_i - \bar{x})^2 * n_i$ |
|---------------------------|-------|-------|-------------|-------------------------------------|
| [1, 7) | 4 | 4 | 16 | $(4 - 533/24)^2 * 4 = 1326,1736$ |
| [7, 15) | 11 | 7 | 77 | $(11 - 533/24)^2 * 7 = 879,3872$ |
| [15, 25) | 20 | 5 | 100 | $(20 - 533/24)^2 * 5 = 24,3837$ |
| [25, 60] | 42,5 | 8 | 340 | $(42,5 - 533/24)^2 * 8 = 3294,0139$ |
| $N = 24 \quad \sum = 533$ | | | | $\sum = 5523,9583$ |

La varianza es:

$$S^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 * n_i}{N} = \frac{5523,9583}{24} = 230,1649.$$

Alternativa 2.

Para obtener la varianza construimos las siguientes columnas en la tabla:

| Tiempo (L_{i-1}, L_i) | x_i | n_i | $x_i * n_i$ | $x_i^2 * n_i$ |
|---------------------------|-------|-------|-------------|----------------|
| [1, 7) | 4 | 4 | 16 | 64 |
| [7, 15) | 11 | 7 | 77 | 847 |
| [15, 25) | 20 | 5 | 100 | 2000 |
| [25, 60] | 42,5 | 8 | 340 | 14450 |
| $N = 24 \quad \sum = 533$ | | | | $\sum = 17361$ |

La varianza es:

$$S^2 = \frac{\sum_{i=1}^k x_i^2 * n_i}{N} - \bar{x}^2 = \frac{17361}{24} - \left(\frac{533}{24}\right)^2 = 230,1649.$$

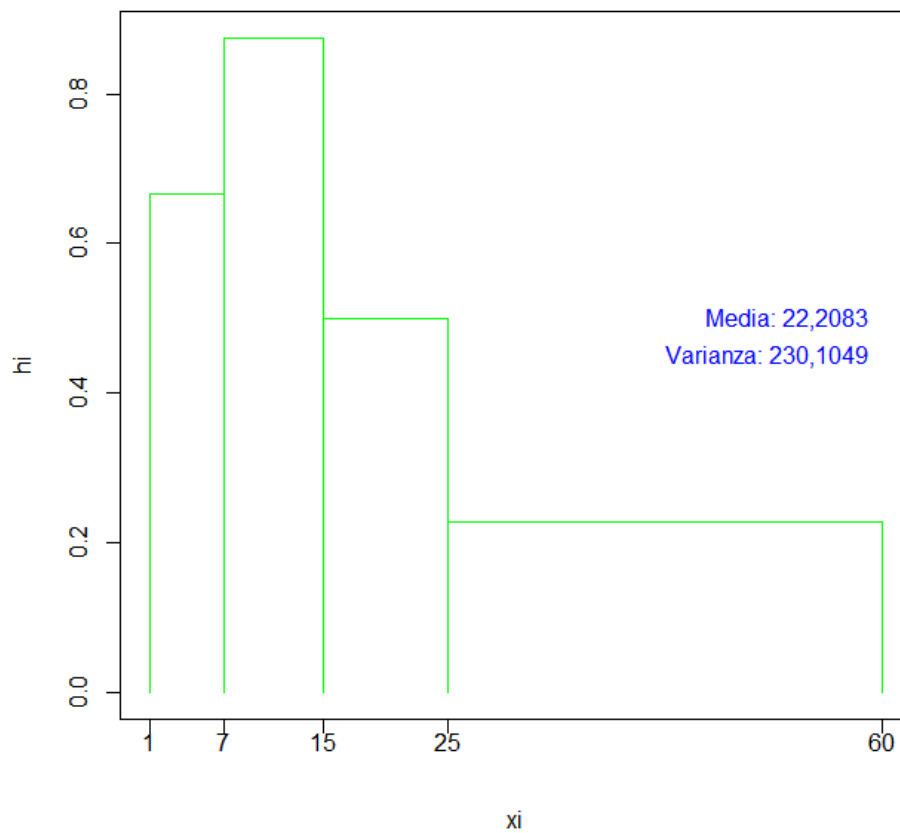
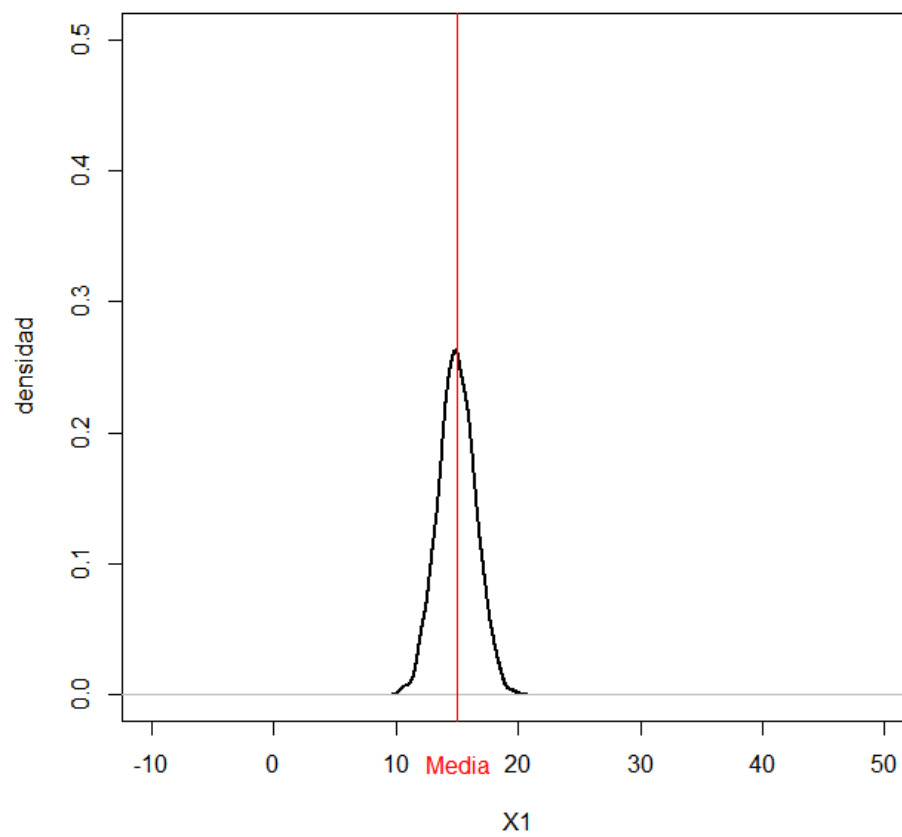
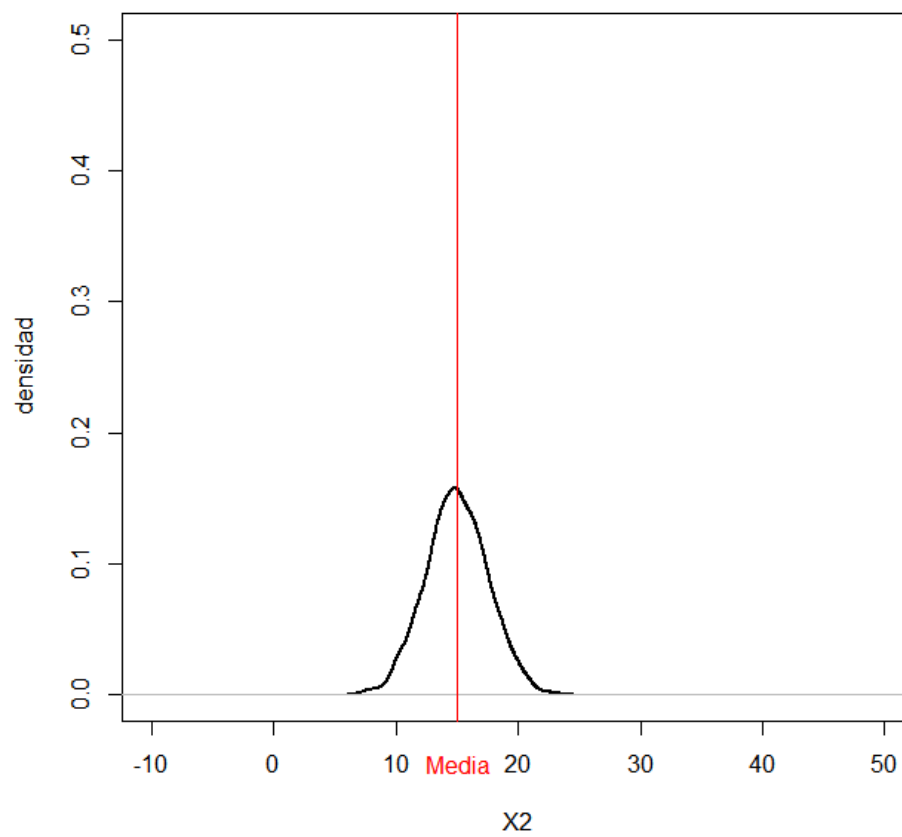


Figura 15.2: Altura asociada al tiempo de vacunación ($h_i = n_i/c_i$)

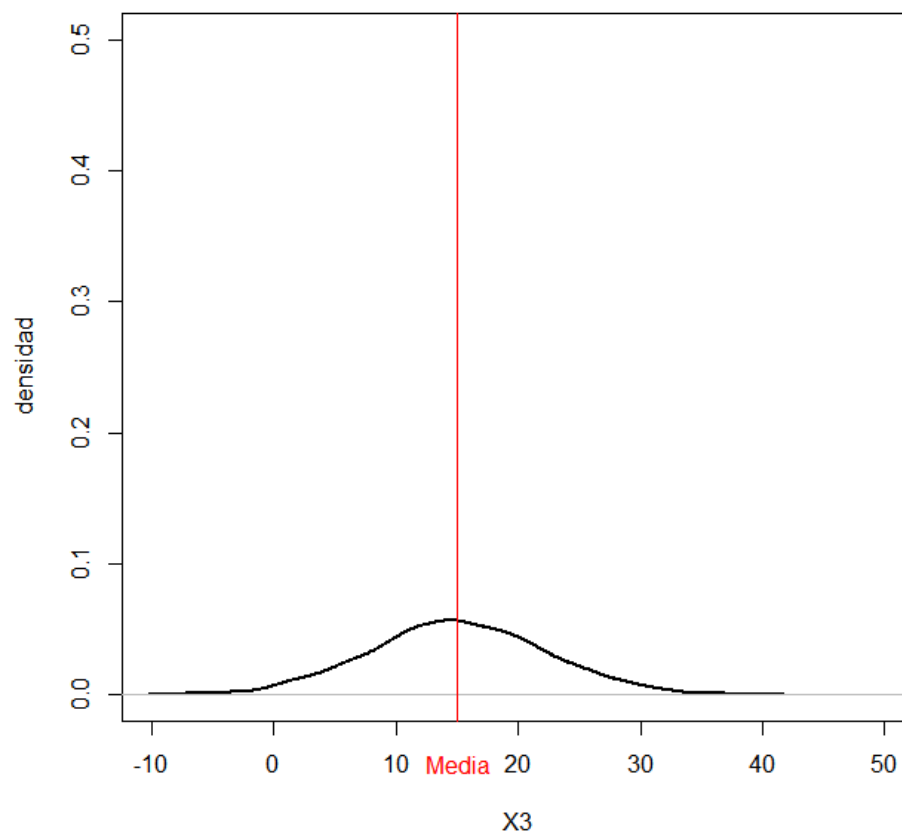
Ejemplo 4. A la vista de los siguientes gráficos, ¿qué variable presenta mayor dispersión?

Figura 15.3: X_1

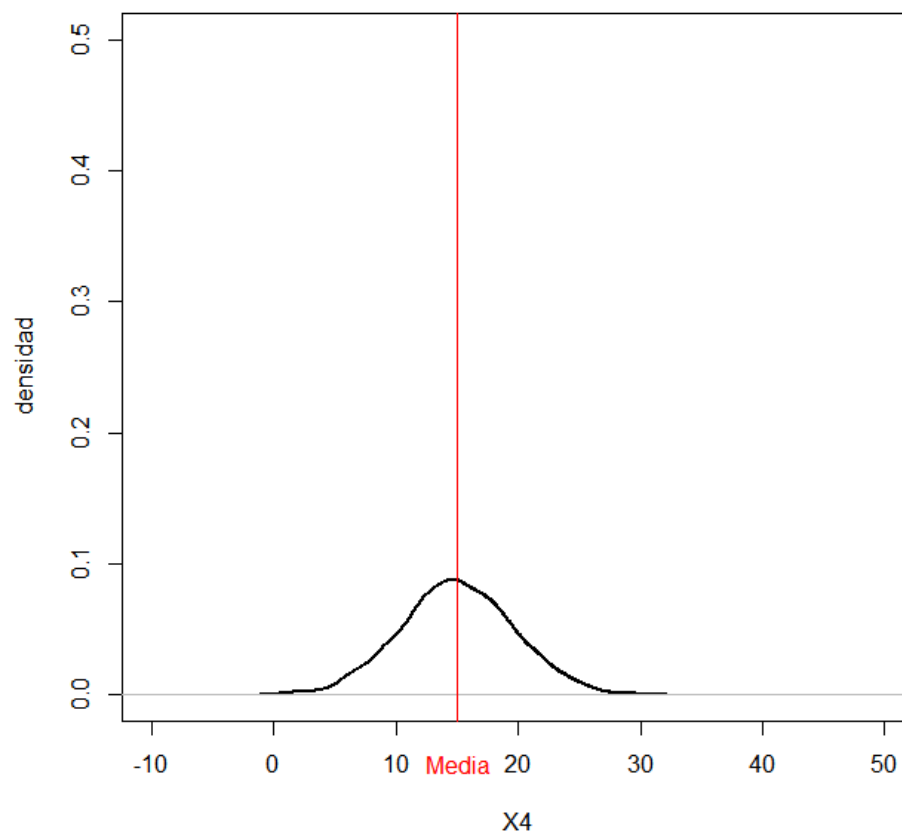
| Variable | Media |
|----------|---------|
| X_1 | 15,0020 |

Figura 15.4: X_2

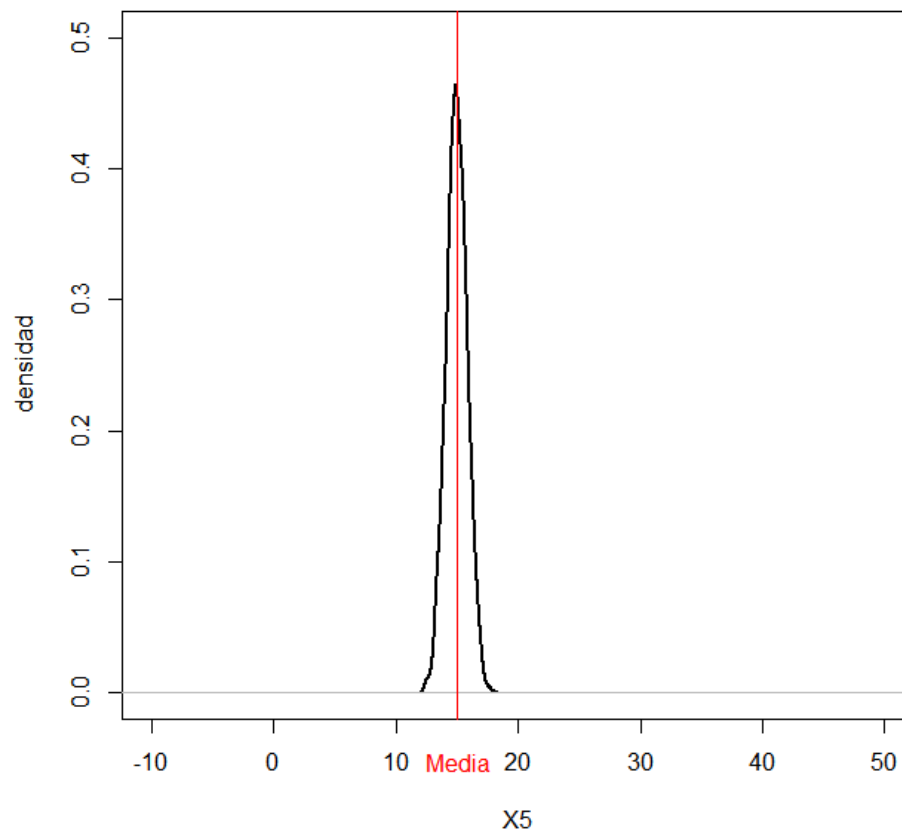
| Variable | Media |
|----------|---------|
| X_2 | 15,0033 |

Figura 15.5: X_3

| Variable | Media |
|----------|---------|
| X_3 | 15,0093 |

Figura 15.6: X_4

| Variable | Media |
|----------|---------|
| X_4 | 15,0060 |

Figura 15.7: X_5

| Variable | Media |
|----------|---------|
| X_5 | 15,0011 |

Solución:

| Variable | Varianza |
|----------|----------|
| X_1 | 2,2725 |
| X_2 | 6,3126 |
| X_3 | 49,490 |
| X_4 | 20,453 |
| X_5 | 0,7297 |

¿Son las distribuciones comparables? ¿Qué se puede hacer en caso negativo? (Nota: véase el coeficiente de variación).

15.3.1 Desarrollo del cuadrado de la fórmula de la varianza

Dada la siguiente expresión

$$S^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 * n_i}{N},$$

si se desarrolla el cuadrado, tenemos la siguiente expresión:

$$\begin{aligned} S^2 &= \frac{\sum_{i=1}^k (x_i^2 - 2 * \bar{x} * x_i + \bar{x}^2) * n_i}{N} = \\ &= \frac{\sum_{i=1}^k x_i^2 * n_i - 2 * \bar{x} \sum_{i=1}^k x_i * n_i + \bar{x}^2 \sum_{i=1}^k n_i}{N} = \\ &= \frac{\sum_{i=1}^k x_i^2 * n_i}{N} - 2 * \bar{x} \frac{\sum_{i=1}^k x_i * n_i}{N} + \bar{x}^2 \frac{\sum_{i=1}^k n_i}{N} = \\ &= \frac{\sum_{i=1}^k x_i^2 * n_i}{N} - 2 * \bar{x}^2 + \bar{x}^2 = \\ &= \frac{\sum_{i=1}^k x_i^2 * n_i}{N} - \bar{x}^2. \end{aligned}$$

Por tanto,

$$S^2 = \frac{\sum_{i=1}^k x_i^2 * n_i}{N} - \bar{x}^2.$$

15.3.2 Cuasivarianza

Relacionado con la varianza, es posible calcular otro estadístico conocido como cuasivarianza. Este estadístico tiene su utilidad en la inferencia estadística. La cuasivarianza se calcula como

$$S_{N-1}^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 * n_i}{N - 1};$$

alternativamente como

$$S_{N-1}^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}.$$

Y su relación con la varianza es la siguiente:

$$S_{N-1}^2 = \frac{N}{N-1} S^2.$$

A partir de la relación anterior, tenemos que $S_{N-1}^2 \geq S^2$.

Ejemplo 5. Retomando los datos del Ejemplo 1, la cuasivarianza es:

$$S_{N-1}^2 = \frac{5}{4} * 8,16 = 10,2.$$

15.4 Desviación típica

Las unidades de la varianza están en unidades al cuadrado de la variable. Por este motivo, como medida de dispersión, también se utiliza la desviación típica, que es la raíz cuadrada positiva de la varianza y se denota por S .

$$S = +\sqrt{S^2}$$

La desviación típica siempre tomará un valor de cero o superior.

Ejemplo 6. Retomando los datos del Ejemplo 1, la desviación típica es:

$$S = \sqrt{8,16} = 2,8566 \text{ grados centígrados.}$$

15.4.1 Cuasidesviación típica

Relacionado con la desviación típica, es posible calcular otro estadístico conocido como cuasidesviación típica. Este estadístico tiene su utilidad en la inferencia estadística. La cuasidesviación típica se calcula como

$$S_{N-1} = +\sqrt{S_{N-1}^2}.$$

Y su relación con la desviación típica es la siguiente:

$$S_{N-1} = +\sqrt{\frac{N}{N-1} S^2}.$$

A partir de la relación anterior, tenemos que $S_{N-1} \geq S$.

Ejemplo 7. Retomando los datos del Ejemplo 1, la cuasidesviación típica es:

$$S_{N-1} = \sqrt{10,2} = 3,1937.$$

15.4.2 Caracterización de una distribución estadística

En la práctica, una distribución estadística se suele caracterizar indicando pocos estadísticos. Principalmente, una medida de posición central y otra medida de dispersión. Los más utilizados son la media y la desviación típica.

Por ejemplo, si utilizamos los datos del Ejemplo 6, podemos caracterizar la temperatura en Rabat durante el mes de julio del siguiente modo:

La temperatura media de Rabat durante el mes de julio es de 19,80 grados centígrados con una desviación típica de 2,86 grados centígrados.

15.5 Otras medidas de dispersión

Existen diversas medidas de dispersión. Además de las presentadas anteriormente, podemos considerar el recorrido intercuartílico o el coeficiente de variación.

15.5.1 Recorrido intercuartílico

El recorrido intercuartílico se calcula como diferencia entre el tercer cuartil y el primer cuartil de la variable. Se denota por RI y su definición viene dada:

$$RI = Q_3 - Q_1.$$

Ejemplo 8. Retomando los datos del Ejemplo 1, podemos calcular el recorrido intercuartílico de la distribución.

Gráficamente, podemos observar que el primer cuartil es 18 grados centígrados y el tercer cuartil es 22 grados centígrados.

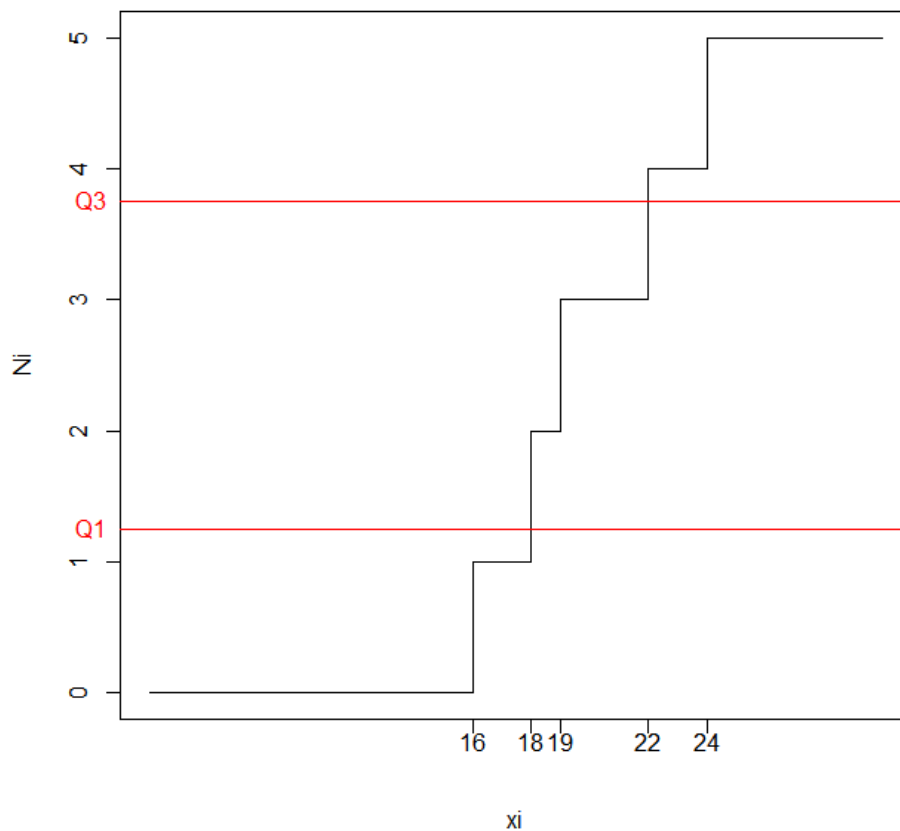


Figura 15.8: Diagrama en escalera de la temperatura en Rabat

Por tanto, el recorrido intercuartílico es:

$$RI = 22 - 18 = 4.$$

Es decir, la diferencia entre el tercer y primer cuartil es de 4 grados centígrados.

15.5.2 Coeficiente de variación

El coeficiente de variación propuesto por Karl Pearson se calcula como el cociente entre la desviación típica y la media en valor absoluto. Es una medida que sirve para comparar el grado de dispersión de dos distribuciones que no vienen medidas en las mismas unidades. Se denota por CV , y su expresión es:

$$CV = \frac{S}{|\bar{x}|}.$$

Valores próximos a cero indican que la distribución presenta baja dispersión (relativa), y por consiguiente, mejor representatividad de la media aritmética.

En la práctica, su valor suele expresarse en porcentaje, por lo que se multiplica por cien. Además, es posible encontrarse en algunos libros la siguiente expresión del coeficiente de variación:

$$CV = \frac{S}{\bar{x}}.$$

En ambas expresiones se supone que la media es distinta de cero.

Ejemplo 9. Retomando los datos del Ejemplo 1, podemos calcular el coeficiente de variación.

El coeficiente de variación es:

$$CV = \frac{\sqrt{8,16}}{19,8} = 0,1443.$$

Recordemos que el coeficiente de variación nos permite comparar la dispersión entre dos o más distribuciones. En este sentido, vamos a ver los siguientes ejemplos.

Ejemplo 10. Retomando los datos del Ejemplo 5, podemos responder a la pregunta formulada en ese ejercicio utilizando el coeficiente de variación.

| Variable | Coeficiente de variación (%) |
|----------|------------------------------|
| X_1 | 10,05 |
| X_2 | 16,75 |
| X_3 | 46,87 |
| X_4 | 30,14 |
| X_5 | 5,69 |

Por tanto, teniendo en cuenta los coeficientes de variación, la variable X_3 es la que presenta mayor dispersión relativa.

Cuando las medias de las distribuciones a comparar sean iguales, la distribución con mayor dispersión relativa será aquella que presente mayor varianza. En el ejemplo anterior, las diferencias entre las medias se situaban en las milésimas y era relativamente fácil identificar la distribución con mayor dispersión relativa sin efectuar cálculos previos.

Ejemplo 11. Dos discográficas han publicado información relativa a la duración de las canciones grabadas durante el verano del año 1992.

La primera discográfica, A, ha reportado que sus grabaciones tienen una duración media de 5 minutos con una varianza de 1,44 minutos cuadrados.

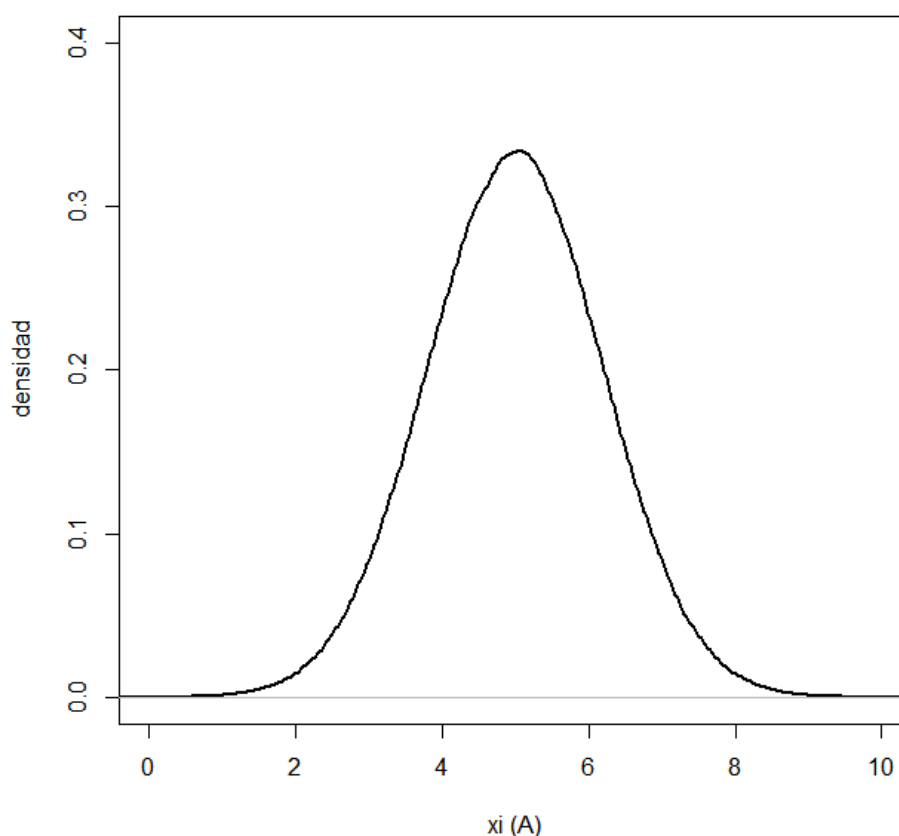


Figura 15.9: Distribución de la duración de las grabaciones de la discográfica A

| Estadístico | Valor de la discográfica A |
|-------------------|----------------------------|
| Media | 5 |
| Varianza | 1,44 |
| Desviación típica | 1,20 |

Mientras que la duración media de la segunda discográfica, B, es de 4 minutos con una varianza de 1,21 minutos cuadrados.

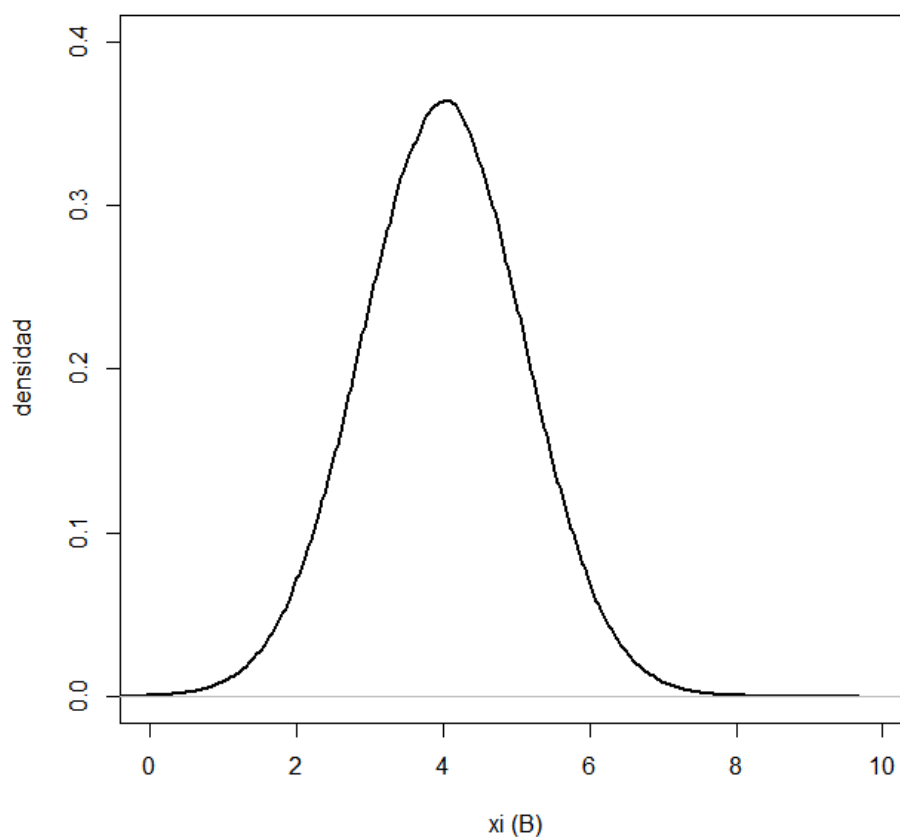


Figura 15.10: Distribución de la duración de las grabaciones de la discográfica B

| Estadístico | Valor de la discográfica B |
|-------------------|----------------------------|
| Media | 4 |
| Varianza | 1,21 |
| Desviación típica | 1,10 |

¿Puedes saber qué discográfica tiene menor dispersión a la vista de los gráficos anteriores?

Al contrario que en el ejemplo anterior, ahora comparar visualmente la dispersión entre las dos discográficas no es tan fácil. Sólo sabemos que la discográfica A tiene mayor dispersión absoluta ($S_A^2 = 1,44 > S_B^2 = 1,21$). Pero esto no nos permite comparar correctamente la dispersión entre las dos distribuciones, ya que en este caso las medias son distintas.

¿Qué discográfica presenta mayor dispersión relativa?

Para responder a la pregunta anterior, calcularemos el coeficiente de variación para las dos distribuciones. El coeficiente de variación es una medida de dispersión relativa que mide el número de veces que la desviación típica de la variable está contenida en la media de la misma. En nuestro caso,

$$CV_A = \frac{S_A}{\bar{x}_A} = \frac{\sqrt{1,44}}{5} = 0,24$$

$$CV_B = \frac{S_B}{\bar{x}_B} = \frac{\sqrt{1,21}}{4} = 0,275.$$

Dado que $CV_A < CV_B$, puede concluirse que la discográfica B presenta mayor dispersión relativa.

15.6 Dispersión y transformaciones

En algunas ocasiones, podemos estar trabajando con la variable X , pero también nos puede interesar realizar una transformación de esta variable (Y). Si previamente ya habíamos calculado los estadísticos de la variable X , para el caso de las transformaciones lineales es fácil obtener los estadísticos de la nueva variable transformada Y .

En este apartado vamos a estudiar cómo obtener algunos estadísticos de la variable tras realizar un cambio de origen y/o de escala.

15.6.1 Cambio de origen

Sumar o restar una constante a todos los valores de la variable con la que estamos trabajando puede hacer que se modifiquen los estadísticos de la varianza, desviación típica, coeficiente de variación y recorrido de la siguiente forma:

| Estadístico | Cambio de origen $Y = X + a$ | Cambio de origen $Y = X - a$ |
|--------------------------|--|--|
| Varianza | $S_Y^2 = S_X^2$ No cambia | $S_Y^2 = S_X^2$ No cambia |
| Desviación típica | $S_Y = S_X$ No cambia | $S_Y = S_X$ No cambia |
| Coeficiente de variación | $CV_Y = \frac{S_X}{ \bar{x}+a }$ Cambia | $CV_Y = \frac{S_X}{ \bar{x}-a }$ Cambia |
| Recorrido | $R_e(Y) = R_e(X)$ No cambia | $R_e(Y) = R_e(X)$ No cambia |

Aquí la constante $a \in \mathbf{R} - \{0\}$.

Ejemplo 12. Supongamos que a una variable estadística X se le ha sumado 7 unidades a cada uno de los valores. Es decir, $Y = X + 7$.

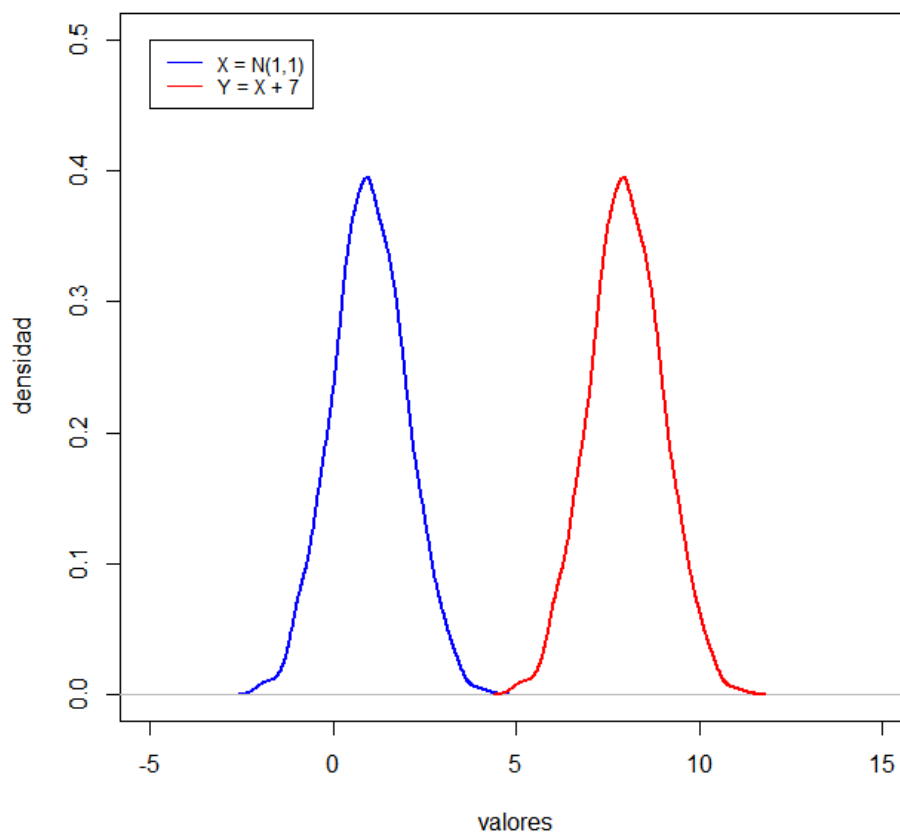


Figura 15.11: Distribuciones de X e Y

| Información de la variable X | |
|--------------------------------|---------------|
| Distribución | <i>Normal</i> |
| Media | 1 |
| Varianza | 1 |
| Mínimo | -2,027042 |
| Máximo | 4,212548 |
| N | 2000 |

Los estadísticos de la nueva variable quedan de la siguiente forma:

| Estadístico | Cambio de origen $Y = X + 7$ |
|--------------------------|--|
| Varianza | $S_Y^2 = 1; S_X^2 = 1$ No cambia |
| Desviación típica | $S_Y = 1; S_X = 1$ No cambia |
| Coeficiente de variación | $CV_Y = \frac{1}{1+7} = 0,125; CV_X = \frac{1}{1} = 1$ Cambia |
| Recorrido | $R_e(Y) = 6,2396; R_e(X) = 6,2396$ No cambia |

15.6.2 Cambio de escala

Multiplicar o dividir por constante a todos los valores de la variable con la que estamos trabajando puede hacer que se modifiquen los estadísticos de la varianza, desviación típica, coeficiente de variación y recorrido de la siguiente forma:

| Estadístico | Cambio de escala $Y = b * X$ | Cambio de escala $Y = \frac{X}{b}$ |
|--------------------------|-----------------------------------|---|
| Varianza | $S_Y^2 = b^2 * S_X^2$ Cambia | $S_Y^2 = \frac{S_X^2}{b^2}$ Cambia |
| Desviación típica | $S_Y = b * S_X$ Cambia | $S_Y = \frac{S_X}{ b }$ Cambia |
| Coeficiente de variación | $CV_Y = CV_X$ No cambia | $CV_Y = CV_X$ No cambia |
| Recorrido | $R_e(Y) = b * R_e(X)$ Cambia | $R_e(Y) = \frac{R_e(X)}{ b }$ Cambia |

Aquí la constante $b \in \mathbf{R} - \{0\}$.

Ejemplo 13. Supongamos que se transforma una variable estadística X multiplicándola por 7. Es decir, $Y = 7 * X$.

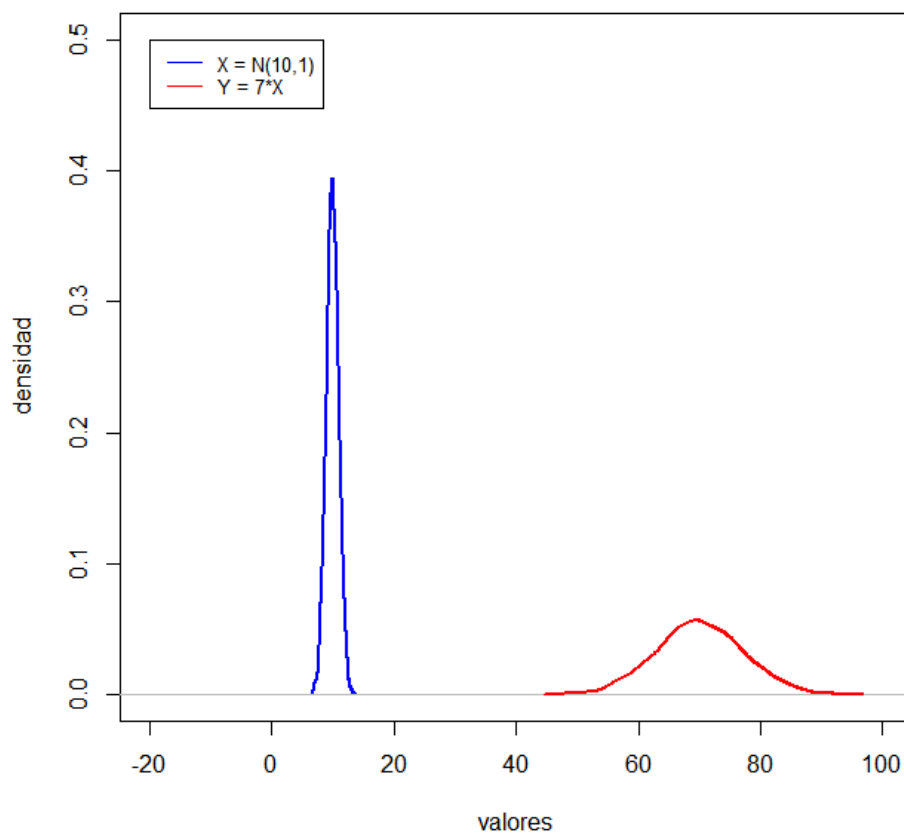


Figura 15.12: Distribuciones de X e Y

| Información de la variable X | |
|------------------------------|---------------|
| Distribución | <i>Normal</i> |
| Media | 10 |
| Varianza | 1 |
| Mínimo | 6,972958 |
| Máximo | 13,212548 |
| N | 2000 |

Los estadísticos de la nueva variable quedan de la siguiente forma:

| Estadístico | Cambio de escala $Y = 7 * X$ |
|--------------------------|---|
| Varianza | $S_Y^2 = 7^2 * 1 = 49; S_X^2 = 1$ |
| | Cambia |
| Desviación típica | $S_Y = 7 * 1 = 7; S_X = 1$ |
| | Cambia |
| Coeficiente de variación | $CV_Y = 0,1; CV_X = 0,1$ |
| | No cambia |
| Recorrido | $R_e(Y) = 7 * 6,23959 = 43,67713; R_e(X) = 6,23959$ |
| | Cambia |

15.6.3 Cambio de origen y escala

Sumar o restar y multiplicar o dividir por constante a todos los valores de la variable con la que estamos trabajando hace que se modifiquen los estadísticos de la varianza, desviación típica, coeficiente de variación y recorrido de la siguiente forma:

| Estadístico | Cambio de escala y origen $Y = b * X + a$ |
|--------------------------|--|
| Varianza | $S_Y^2 = b^2 * S_X^2$ |
| | Cambia |
| Desviación típica | $S_Y = b * S_X$ |
| | Cambia |
| Coeficiente de variación | $CV_Y = \frac{ b * S_X}{ b * \bar{x} + a }$ |
| | Cambia |
| Recorrido | $R_e(Y) = b * R_e(X)$ |
| | Cambia |

Aquí las constantes a y $b \in \mathbf{R} - \{0\}$.

Ejemplo 14. Supongamos que se transforma una variable estadística X sumándola y multiplicándola por 7. Es decir, $Y = 7 * X + 7$.

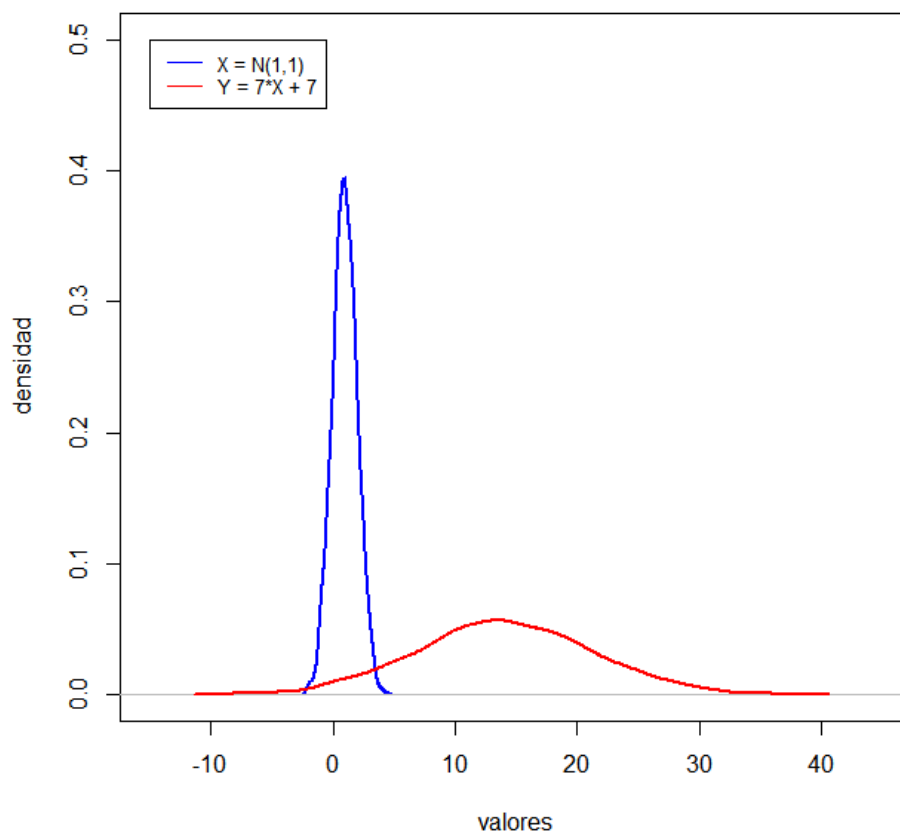


Figura 15.13: Distribuciones de X e Y

| Información de la variable X | |
|------------------------------|---------------|
| Distribución | <i>Normal</i> |
| Media | 1 |
| Varianza | 1 |
| Mínimo | -2,027042 |
| Máximo | 4,212548 |
| N | 2000 |

Los estadísticos de la nueva variable quedan de la siguiente forma:

| Estadístico | Cambio de escala $Y = 7 * X + 7$ |
|--------------------------|---|
| Varianza | $S_Y^2 = 7^2 * 1 = 49; S_X^2 = 1$ |
| | Cambia |
| Desviación típica | $S_Y = 7 * 1 = 7; S_X = 1$ |
| | Cambia |
| Coeficiente de variación | $CV_Y = 0,5; CV_X = 0,1$ |
| | Cambia |
| Recorrido | $R_e(Y) = 7 * 6,23959 = 43,67713; R_e(X) = 6,23959$ |
| | Cambia |

El caso del recorrido intercuartílico es similar al del rango o recorrido. Los valores del primer y tercer cuartil se ven afectados por las transformaciones lineales (origen y escala), por lo que el recorrido intercuartílico se ve modificado de forma idéntica al rango. Es decir, ante un cambio de escala, el recorrido intercuartílico de la nueva variable transformada (Y) será igual al producto (o cociente) del valor absoluto de la constante (b) por el recorrido intercuartílico de la variable sin transformar (X).

15.6.4 Variable tipificada

Una variable tipificada es una nueva variable que se obtiene mediante una transformación lineal del tipo

$$Z = \frac{X - \bar{x}}{S}$$

De esta forma se consigue eliminar la influencia de la media y la desviación típica sobre la variable X . Este tipo de tipificación se conoce como tipificación estándar y la nueva variable tipificada se denomina Z . La nueva variable así obtenida tendrá media 0 y desviación típica 1.

Por tanto, para comparar dos distribuciones estadísticas se puede recurrir a la tipificación de variables. Particularmente, es posible comparar valores de poblaciones diferentes, o variables que miden conceptos distintos. La comparación de dos valores tipificados se hace en términos relativos, ya que los valores tipificados indican la distancia a la media (del valor original) en unidades de desviaciones típicas.

Ejemplo 15. Un profesor que trabaja en la ciudad A tiene un salario anual de 45.000€; y otro que trabaja en la ciudad B cobra 70.000€ anuales. Si las medias y desviaciones típicas son: 21.900€ y 11.000€ en la ciudad A; 52.500€ y 9.000€ en la ciudad B, respectivamente. ¿Qué profesor tiene una mejor posición en términos salariales?

Calculamos los valores tipificados:

$$z_A = \frac{45000 - 21900}{11000} = 2,1$$

$$z_B = \frac{70000 - 52500}{9000} = 1,94$$

Dado que $z_A > z_B$, en términos relativos, el profesor de la ciudad A tiene una mejor posición salarial en comparación con la posición que ocupa el otro profesor de la ciudad B.

Problema propuesto. Un economista desea analizar la cotización de un grupo de empresas españolas. Para ello, se ha fijado en el precio de la acción en euros (X) y el efectivo en miles de euros (Y) en una sesión del Mercado Continuo. Los datos se presentan en la siguiente tabla:

| X | n_i | Y | n_j |
|--------|-------|------------------|-------|
| 0,54 | 1 | [2800, 5000) | 3 |
| 30,7 | 6 | [5000, 10000) | 7 |
| 78,8 | 4 | [10000, 15000) | 6 |
| 153,5 | 8 | [15000, 25000) | 4 |
| 244,5 | 9 | [25000, 40000) | 7 |
| 559,4 | 5 | [40000, 70000) | 3 |
| 1300,2 | 2 | [70000, 100000) | 3 |
| | | [100000, 170000] | 2 |

Se pide:

- ¿Cuál es la diferencia entre la acción más cara y la más barata?
- ¿Cuál es el recorrido intercuartílico del efectivo?
- Estudia la variabilidad del efectivo. Nota: calcula la varianza y la desviación típica.
- Un inversor quiere invertir en una cartera de acciones si su coeficiente de variación es inferior a 0,25. Ante una cartera como la de tus datos, ¿qué hará el inversor?

Por último, se recomienda la lectura de otros libros de varios autores para ampliar los conocimientos respecto de la materia estudiada (**bachero2006estadística; pena2001fundamentos**).